Electrical impedance tomography applied to body-support interface pressure measurement

Fulton, William Sean

Award date:
1995

Awarding institution:
University of Bath

ELECTRICAL IMPEDANCE TOMOGRAPHY APPLIED TO BODY-SUPPORT INTERFACE PRESSURE MEASUREMENT

submitted by William Sean Fulton BEng
for the degree of PhD
of the University of Bath
1995

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This thesis focuses on body-support interface pressure measurement. The intended applications for interface pressure measurement are in the medical field, to study the pressure distributions of seated wheelchair patients or bed-ridden patients. In certain circumstances, these patients will be prone to pressure sore development. The required pressure measurement data will enable medical physicians to study the pressure distributions that lead to pressure sores. The data will also enable the assessment of the different prophylactic devices that alleviate and spread pressure. A system could also be placed in each hospital ward to estimate the probability of pressure sore formation in newly admitted patients.

This proposed pressure measurement system is unique in that it measures pressure over a continuum. Current pressure measurement systems are limited in that they only give pressure readings at certain points. The proposed system needs a thin carbon fibre felt mat placed between the body and supporting medium. The mat exhibits a change in resistivity at points at which a load is applied. Electrical Impedance Tomography (EIT) is a medical imaging technique that can be applied to reconstruct the resistivity distribution by applying currents to the mat. The usual applications for EIT are in imaging body organs, bones, tissues, etc. where access to the centre of the body is prohibitive. In the case of a resistive mat, access to the centre of the resistivity distribution is available. Central access reduces the effects of data measurement error on image accuracy. Additionally, the largest errors of three-dimensional current flow and variable electrode positioning are removed and reduced respectively.

The thesis concentrates on the modelling of the pressure sensitive mat and assessing the viability of applying EIT to pressure measurement. The computer model discretises the domain through the use of the finite element method. The finite element model is based on square bilinear elements and their choice is discussed. The model chosen allows the use of a nested dissection re-ordering algorithm. Combining this sparse matrix algorithm with the sparse vector method, the time required to reconstruct the pressure distribution is improved. Other techniques to reduce reconstruction time are also detailed; these are primarily based around the Jacobian matrix formation, a matrix required by the reconstruction algorithm.

EIT APPLIED TO BODY-SUPPORT INTERFACE PRESSURE MEASUREMENT

Simulations using medical pressure profiles, such as the foot and buttocks, are reconstructed by various least squares algorithms. The Gauss-Newton algorithm regularised by second finite differences is chosen as the most suitable for reconstruction. Reconstruction accuracy is examined and it is shown that image improvements are obtained when attaching electrodes to the centre of the resistive mat. Spectral analysis helps to determine the most pragmatic electrode configurations for data gathering. Using practical electrode configurations, the reconstructed profiles indicate great promise for implementing a real-world EIT system for interface pressure measurement.

Table of Contents

List of Abbreviations

APT	Applied Potential Tomography
CPU	Central Processing Unit
EIT	Electrical Impedance Tomography
FDM	Finite Difference Method
FEM	Finite Element Method
FB	Fast Back
FF	Fast Forward
G-N	Gauss-Newton
IND	Irregular Nested Dissection
MD	Minimum Degree
NAG	Numerical Algorithms Group Ltd
NHS	National Health Service
RAM	Random Access Memory
RND	Regular Nested Dissection
SVD	Singular Value Decomposition

1. Introduction

The aim of this section is to give an overview of pressure measurement and pressure sores. The prevention of them is the target application for the use of a body-support interface pressure measurement system. Other medical uses for such a system are mentioned. Current pressure measurement techniques are analysed and the requirements for a good pressure measurement system are examined. The section is concluded with the proposed pressure measurement system.

RA Knight started the research in this field and the research contained within this thesis is a continuation of the work he initiated. Part of this first chapter is, in effect, an overview of the first two sections of his thesis[1]. This chapter also includes further research into newly published work.

1.1. Pressure Sores

1.1.1. Introduction to Pressure Sores

A pressure sore is an ulceration of the skin and/or deeper tissues due to unrelieved pressure, shear force(s), and/or frictional force(s). Uninterrupted pressure on the skin above the capillary pressure of 32mmHg reduces blood flow to the area under pressure and so produces local ischaemia, which if prolonged will cause an ulcer. Modern research has analysed this further and shown that skin blood flow does not cease until the applied pressure is above diastolic pressure[2]. When an applied pressure is greater than capillary pressure, skin blood flow continues, but at a reduced level. With limited or no blood flow, tissue survives for a short time. It is of paramount importance to relieve pressure early in order to prevent pressure sore formation.

Pressure sores are a nasty infliction occuring in tens of thousands of patients in Britain. Pressure sore patients demand difficult and labour intensive nursing care, and suffer degrees of physical and mental discomfort which cannot be measured in purely economic terms. The most effective means of reducing the impact of pressure sores is by the prevention of them. Improved methods of treatment would at the most only have a marginal effect[3] on financial costs. Treatment should thus be concentrated on pressure sore prevention.

Surveys of hospital and community populations have shown that pressure sores account for 25% of the cost of treatment and maintenance of patients so afflicted. Indeed, pressure sores are a great financial burden on the NHS. Exton-Smith[3] estimated the cost to be £420 million p.a. To this must also be added the cost of treating sores outside hospital. In the USA, the medical costs associated with pressure sores are estimated to be US$2 billion[4]. 70% to 90% of sores occur in the over 75 age group[5] and with the ever increasing elderly population and life expectancy, the number of potential pressure sores sufferers and consequent cost is unfortunately also on the increase.

1.1.2. Causes of Pressure Sore Formation and the Persons They Affect.

Impairment of the body's micro circulation due to pressure, automatically invokes certain biological and biochemical responses, one of which generates a neurological response of discomfort. In normal circumstances, i.e. in healthy people, this provokes a voluntary movement of the body or limb to redistribute the load, thereby improving micro circulation before cell damage is irreversible. However, in abnormal circumstances this biofeedback signal cannot initiate any remedial action and irreversible cell damage will ultimately occur. Examples of such abnormal circumstances, where pressure sore formation is common, are in the aged, in operating rooms, in people who are physically injured and in paraplegics. Operating theatre tables are very firm and if the length of an operation is more than two hours, a pressure sore is likely to form. Movement is often not possible for physically injured patients, for example, those with burns or paralysis. Pressure sores are common amongst people who are permanently paralysed, especially those in wheelchairs who do not regularly sit up to relieve pressure on the buttocks. Until the widespread use of antibiotics, pressure sores were not such a problem in the elderly as they succumbed to the initial infection and did not have time to develop them. However, the formation of a pressure sore is often the cause of death in the elderly now. In fact, any patient with an illness impairing the natural protective mechanisms of movement is susceptible, especially if their physical condition is poor. Other examples include mental apathy, comas, clouding of consciousness, anaesthesia, senility, malnutrition, failing blood circulation and skin diseases; e.g. leprosy.

The most important single factor responsible for the development of sores is sustained pressure on the tissues which leads to obstruction of blood vessels, ischaemia and necrosis of tissue[3]. The areas most at risk are mainly in the bony prominences, where the insufficiency of underlying tissue is unable to redistribute pressure. Pressure sores on

anaesthetised patients occur during short operations on hard operating tables and in immobile patients on soft beds who don't move for a few hours. Versluyen[6] in his study, attributed the formation of pressure sores to the long periods during which the patients were immobilised on high pressure surfaces in casualty, wards and theatre before the repair of a fracture or the restoration of weight-bearing function. Pressure of sufficiently long duration to cause tissue damage, rarely occurs in healthy individuals, since abnormally high pressures are only sustained for short periods. Even in sleep, the limits of tolerance to pressure are seldom exceeded since position changes occur 20-60 times during the night. For example, in leg muscle, the threshold of 100mmHg for two hours must be reached to produce definite microscopical changes, which ultimately leads to pressure sores; the effects are similar to those produced by 600mmHg for one hour. Exton-Smith[3] concludes that high pressure maintained for short periods is damaging, but not as damaging as low pressure for long periods.

The various factors which cause pressure sores can be broken into two groups: 1) intrinsic factors and 2) extrinsic factors. Intrinsic factors are concerned with aspects of the physical or medical condition of the patient and extrinsic factors are derived from the environment of the patient. These two factors are discussed further by Knight[1].

1.1.3. Prevention of Pressure Sores

Identifying those patients at risk and removing the causes of tissue necrosis along with subsequent patient monitoring is the ultimate goal. Regular patient turning is the ideal solution to pressure sore prevention, but many underfunded and understaffed wards simply do not have the facilities to devote time to this task. Exton-Smith[3] concludes that it is now unrealistic to expect nurses to practise two-hourly turning, except in rare instances, and scarce nursing skills should now be deployed for the recognition of high-risk patients who require the use of specialised equipment for the prevention of pressure sores.

Although patient turning is the ideal method of prevention, this is not always possible. In these cases and to ease nurses' workload, special airbeds, waterbeds, mattresses, cushions, sheepskins etc. are used to distribute pressure as evenly as possible. Pressure distributing devices are also needed on some operating tables and special cushions in the case of chair bound people.

There are many other techniques used in hospitals for the prevention of pressure sores; these are examined in more detail by Knight[1]. He noted that as a result of the increasing recognition of the cost of pressure sores, more and more hospitals are developing pressure sore prevention policies. He looked at various patient support systems in use at hospitals that reduce constant high pressures on the body. These include devices that mould to the body to redistribute pressure and those that move the areas of high pressure to different points under the body.

Clinical use suggests that pressure relieving support systems are quite effective, however, their expense excludes them from general hospital use and many of the designs and constructions inhibit patient handling. If these methods of prevention are not available, Knight[1] emphasised the importance that patients are turned regularly as a single failure to do so will subject a patient to the formation of a pressure sore.

To summarise, the management of pressure sores and their prevention can be seen to fall into three broad categories. Firstly and most importantly, is the removal of pressure. Secondly, there is a necessity to reduce the factors leading to pressure sores and finally to implement a wound care programme to treat those who do develop pressure sores. The costs that pressure sores incur on hospitals has been shown to be high, despite the proliferation of beds and support surfaces designed to prevent them occurring. The magnitude of the pressure sore problem appears to be so large and all pervading that any device or technique that leads to their prevention and to any saving of the costs involved would be beneficial to hospitals, but even more so to the poor sufferer of such sores.

1.2. Pressure Measurement.

The measurement of pressure in the clinical environment has not yet satisfactorily been completed. Much of the research into pressure sores has provided inconclusive or contradictory evidence for sore formation. Part of the reason for this is the inaccurate measurement of pressure, and it is this variable which is ultimately the underlying factor to sore formation. The following section gives an insight into previous pressure measurement systems.

1.2.1. Previous Pressure Measurement Systems

Knight[1] in his thesis, looked into previous measurement systems, both old and new. In all the measurements systems he researched, limitations were found. Grant[7] too has reviewed current interface pressure measurement systems that are of use. Both Knight

and Grant have similar conclusions in the limitations of these systems. These are that the readings have somewhat limited information content, non-repeatability and are not suitable for the environment needed e.g. the transducers are rigid. In the practical environment, according to Grant, interpretation of the readings to gain any sort of reliable information is a skilled task and an ideal solution still does not exist. It is in fact LJ Grant, Head of Medical Physics, Wessex Regional Health Authority, who brought to our attention, the need for a good pressure measurement technique and it is his department who inspired our research into this area.

Present pressure measurement systems are based on individual sensors. These sensors, due to their large physical size, measure point pressures. When placed in a matrix, they are known to miss out peak pressures, as well as interfere with the real pressure distribution. The required spacing of the individual sensors depends upon the radius of curvature of the anatomical site being studied. These sensors are also limited in their point pressure accuracy.

One type of commercially available pressure sensor is the Force Sensitive Resistor manufactured by Interlink Electronics. These sensors consist of a polymer sheet with a layer of their proprietary pressure sensitive ink screen printed onto it. Each sensor is flexible and can be manufactured to sizes of 4mm in diameter or larger and is just 10μm thick. They exhibit a log pressure-resistance characteristic. These sensors have been used in various recent applications and research results have been produced from these sensors, for example a 16×16 array in robot tactile sensing[8] and in an 8×8 pressure mat targeted at pressure sore prevention[9]. Colleagues at Odstock Hospital, Salisbury have used the sensors as footswitches, but found them to be unreliable after time. The sensors exhibited unworkable hysteresis and non-repeatability. One cause of this was the wearing away of the pressure sensitive ink from shear forces. The sensors also do not exhibit repeatable characteristics if they are bent. They are prohibitively expensive, particularly so if they must be replaced after wearing out.

Two other new pressure-sensitive devices have just recently been documented. The first[10] consists of an electropneumatic flat disc shaped sensor cell connected by a tube to a piezo-resistive pressure-sensitive device. The sensor cell is filled with vegetable cooking oil, transmitting the pressure from the patient to the piezo-resistive transducer. The results are quite promising as far as accuracy is concerned (pressures within 0.5mmHg); furthermore, it gives a continuous electrical output, a failure in previous

systems. The system is cheap, which is very important, and is constructed of commercially available parts. This device shows very little hysteresis, minimal time and temperature drift and good linearity. However, as is typical of individual pressure sensors, it is quite large in diameter - 28mm. A pipe 3mm in diameter, is connected to the sensor and this interferes with the pressure profile of the subject being measured. The designers suggest, but never tested, using the sensors in a matrix configuration for multiple pressure measurements. This would have the limitations already mentioned.

The second recently published device is a pressure-sensitive mat for measuring contact pressure distributions of patients lying on hospital beds[11]. The system includes a flexible pressure sensitive mat, electronics to activate the mat, a small computer to process the data and a colour video display. The mat incorporates 1536 discrete capacitive nodes giving a spatial resolution of 25mm. The pressure-sensitive mat itself, includes two orthogonal arrays of ribbon-like conductors, composed of silver-coated nylon fabric, which are separated by insulating open-cell foam rubber. The system monitors the node capacitance between selected pairs of horizontal and vertical conductors. The node pressure is determined from the capacitance, which increases with the application of pressure. The mat is, unfortunately, liable to interfere with the pressure distribution at the interface due to its thickness - 5mm. It is one of the few large pressure profile measuring mats that have been developed, but the designers admit that a variety of problems remain to be solved before pressure-sensitive mats based upon capacitive transduction can be clinically useful tools. These problems are mainly to do with the hysteresis of the foam rubber and its variance in capacitance with ambient humidity.

The above design measures point pressures over a mat by calculating the change in capacitance when a force is applied. The same technique could be applied, but replacing the capacitive sheet with a conductive rubber which changes its resistivity with an applied load. The technique of Babb[11] could then be used to measure node resistance. This would involve the bulk of the sheet comprising strips of an insulator alternating with strips of force dependent resistive material, as shown in Figure 1.1. The force dependent material would then require a very thin conductive layer embedded into one surface. Access to a node would be completed by embedding perpendicular conductive strips into the opposite sheet surface. This method of measuring pressure over an area was rejected for the following reasons:

- The manufacturing process of the material would be complicated and thus expensive.
- The pressure map would not be a continuous one as only the pressure at nodes is obtained.
- A higher number of electrical connections to the mat would be needed to obtain the same resolution produced by the Electrical Impedance Tomography technique proposed in this thesis.

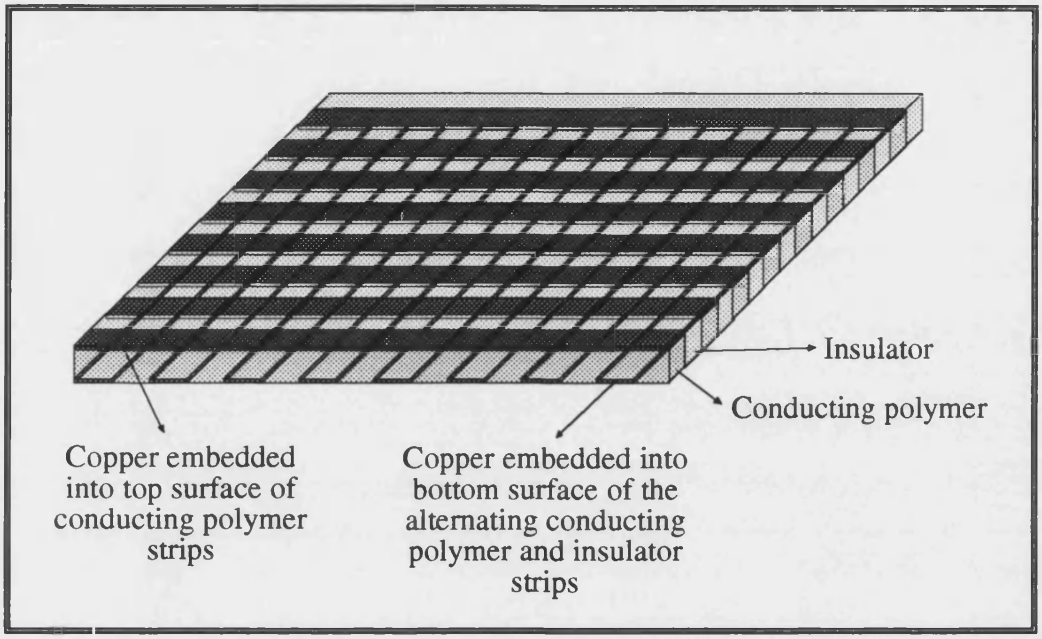


Figure 1.1 Possible pressure measuring mat construction showing alternating strips of insulators and pressure sensitive conducting polymers.

For the ideal pressure measurement of a patient on a support surface, an entire pressure profile of the area of interest needs to be mapped over a length of time. The aforementioned reasons leads Grant[7] to conclude that individual transducers, even if they are very accurate, do not fulfil the required needs in a hospital environment. What is needed is a thin, flexible mat to map the entire region of interest.

1.3. Medical Uses of Pressure Measurement.

The first use of a pressure measuring system, would be for medical physicians to use as a research tool to evaluate specialised beds and wheelchairs. Should the system be priced low enough, it would be used in most hospital wards throughout the country to monitor those patients who are most prone to a sore being developed.

There are of course numerous other applications for such a device. Those that are in the medical field are summarised as follows:

- Individual wheelchair cushion design. These cushions need to spread the pressure, again to prevent pressure sores from forming.
- The design and adjustment of orthopaedic footwear and objective pressure measurement of anaesthetic feet.
- Load distribution in the analysis of gait.
- Balance aid and posture analysis obtained from pressure distributions of both feet.
- Device to ensure teeth meet correctly during jaw operations.
- Movement detector to monitor breathing of children for cot death prevention.
- Training paraplegic patients to stand, with the aid of muscle stimulation, by measuring their centre of gravity.

1.4. Requirements for the Pressure Measurement System

Garber[4] gives a conclusive need for such a mat as an aid to pressure sore prevention:

"...In order to maximise the impact of technology in reducing the occurrence of pressure sores, it is necessary that users of technology - clinicians and engineers alike - reorient their efforts to develop instruments that can monitor tissue viability under a seated person or a person who is lying in bed. The technology must be able to detect when the tissue is in danger of dying while the process is still reversible so that appropriate measures can be initiated."

Knight[1] summarises a list of requirements which would fulfil the aims of Garber[4]; these are reproduced below.

- It must be flexible enough so as not to interfere with the body-support surface interface pressures. The ideal position for the sensor is clearly at the interface since this is the best location at which to observe the pressure distribution over the surface of the skin. If the sensor is placed beneath the supporting medium the interface pressures may be reduced or nullified by the effects of the supporting medium. Moreover, Reddy[12] stated that the ideal interface pressure transducer should support all of the load, to reduce its effects on the interface characteristics.
- It must be able to locate, to a reasonable degree of accuracy, the positions on the surface of the body that are subject to unnaturally high pressures. These generally occur around the bony prominences of a lying or seated patient. Thus, a resolution of

around 2cm, as used by Lindan[13], would appear to be suitable. Typical quantitative values of pressure that must be measured at the interface are given later in this thesis. [These were shown to be in the range 0-150mmHg.]

- If the system is to be used as an aid to the prosthetist, its ability to measure relative pressure distribution is more critical than its ability to measure absolute pressure. The aim of the prosthetist and cushion designer is to provide a support surface that equalises pressure. However, if the sensor system is to be used as a real-time pressure monitoring system, that is, mainly as an aid in the field of pressure sore prevention, then absolute pressure measurement is essential. The values of pressure, along with its duration, could then be processed by a pressure sore prevention algorithm, designed to alert nursing staff, or the patients themselves, to the high pressure areas.
- The system should be microcomputer controlled. The data extraction and processing can then be performed automatically, allowing nursing staff to attend to other tasks. The provision of data should allow easy and effective interpretation by the nursing staff, the equipment should be portable and be as cheap as possible.
- The system must be fail-safe, conforming to the relevant British Standards, and must have a low failure rate and recalibration requirement. As an aid to maintenance, the system should have test facilities and be based on an easily repairable modular design.
- The sensor itself must be durable, robust and impervious to bodily fluids and excreta.
- More particularly, it must be comfortable to lie on.

It must still be noted that the information obtained by measuring pressure does not completely characterise the body-support interface, but, even with this limitation, pressure measurements are important[14]. Barbenel[14] also states that the clinical usefulness of pressure measurement has been most clearly established and that both pressure and time for which it acts are important variables in assessing the potential for producing tissue damage by any support surface.

1.5. The Proposed EIT Body-Support Pressure Measurement System

The Body-support pressure measurement system is quite simple in its structure. The pressure sensing mat is placed on the support surface (bed or wheelchair) and a number of wires connects it to the interface electronics. The interface electronics is connected to a computer with monitor for calculation and display of pressure maps.

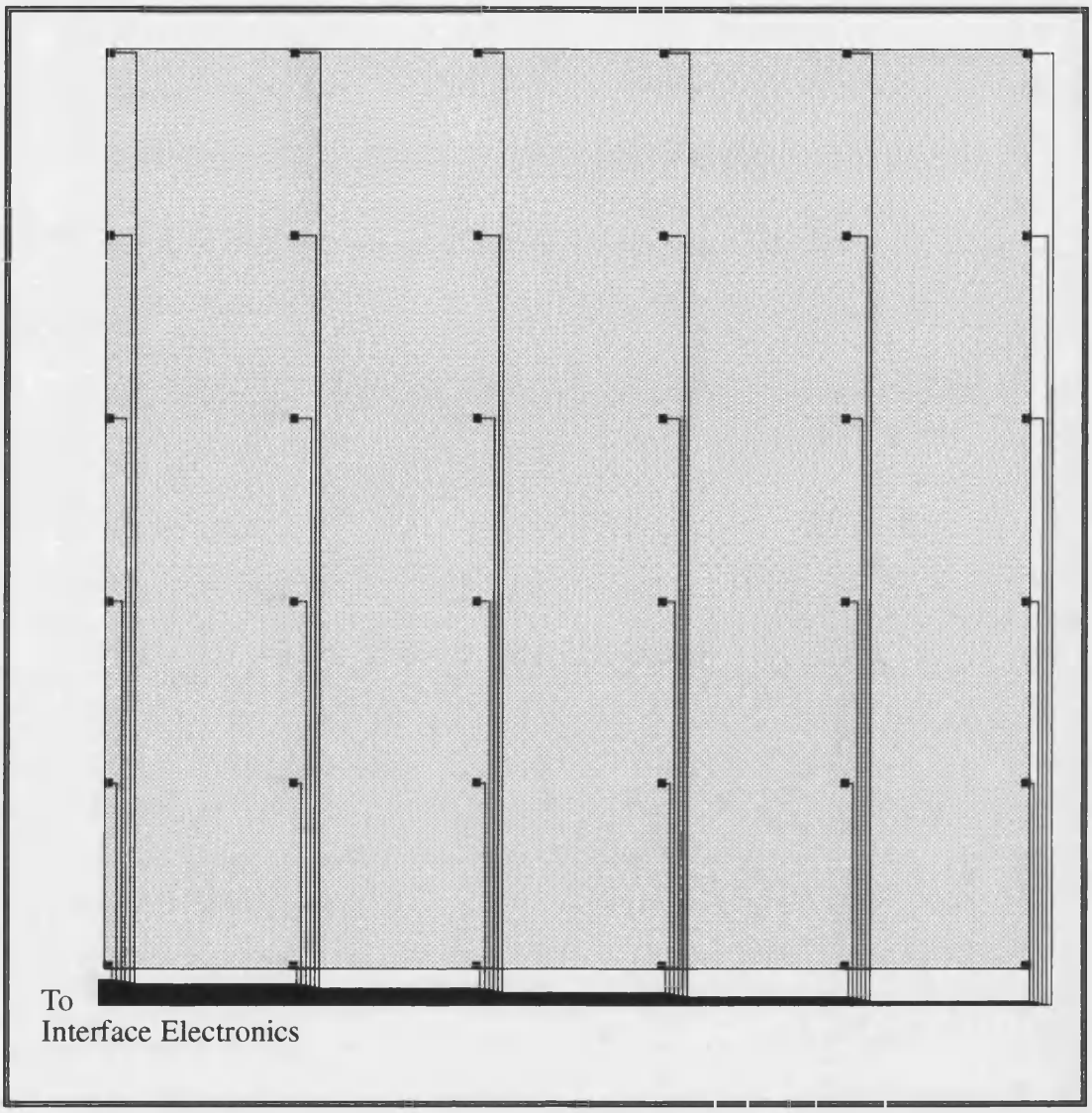


Figure 1.2 Top view of the mat showing electrical connections to 36 point electrodes.

The mat must be very thin to reduce interference at the interface of the body and support surface. The ideal sensor mat is made from a sheet of conductive material, which exhibits known characteristics relating its electrical resistivity to the pressure exerted on it. Thus, by reconstructing the two-dimensional resistivity distribution on the conductive sheet, a distribution or "picture" of the applied pressure between the body and the supporting surface is known. Proposed mat materials are examined in Chapter 2 and conform to the necessary requirements.

To reconstruct the resistivity profile, electric currents must be applied to the mat and the resulting electric voltages need to be measured. To enable this a multiplicity of electric contacts to the mat must be made. Figure 1.2 illustrates the electrode positioning and wiring. Connections are made at the periphery as well as the central area of the mat. Those that are connected to the interior must be electrically isolated from the conductive mat, except at the end of the wire connector (electrode), where they are connected to the mat at a point. The internal wire connectors must be thin, preferably flat, in order to adhere to the requirement of minimal disturbance to the real pressure distribution.

The interface electronics must consist of high precision current drivers, high precision voltage measurement and a digital interface to the computer for control purposes and the exchange of data. The primary function of the interface electronics is thus to collect data from the mat and send it to the computer.

All the calculations required to produce the resistivity profile of the mat and hence pressure distribution need to be performed on a suitably powerful computer. The computer enables a good interface to the user by use of its monitor. The monitor can easily display an excellent visual description of the pressure distribution as well as an analysis of the distribution. This is necessary if the mat is used as a research tool. However, if it is being used as a patient monitoring system, with the correct pressure-time analysis, all that is needed is for an audible or visual alarm to be sounded when necessary. In this case, a dedicated microprocessor system could replace the computer. The calculations required to obtain the desired pressure distribution are based on a reconstruction technique called Electrical Impedance Tomography, and it is this technique that the bulk of this thesis addresses.

The following article by Andrew Watson was published in New Scientist, page 19, on 14 December 1991 and is reproduced here with the kind permission of New Scientist.

The perplexing puzzle posed by a pile of apples

Andrew Watson, Melbourne

IN A pyramid of apples at the greengrocer's, which apple is being crushed the most? The obvious answer is the one directly under the apex, at the bottom of the pile. But the obvious answer is wrong.

In 1981, Czech scientists J. Schmid and J. Novosad used pressure sensors to measure the forces acting at the base of a pile of particles, of which the pyramid of apples is an everyday example. They discovered an unexpected pattern of forces, which is proving hard to explain.

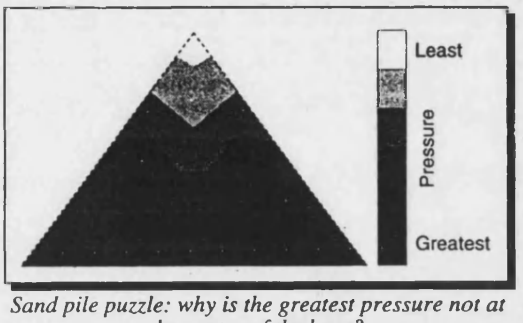
Schmid and Novosad found that the pressure is not greatest at the central point of the base, as might be expected. Instead, a ring of particles some distance from the centre of the base feels the greatest downward pressure.

In fact, at the central point of the base, the pressure shows a local minimum. Its magnitude relative to the maximum pressure felt by particles depends on the size of pile. For the largest pile investigated by the Czech reasearchers, the local minimum was 30 per cent less.

Schmid and Novosad also found that the shear, or sideways, forces showed a minimum at the centre of the base. The forces increased for a while towards the outside of the pile, then tapered off again at the outer edge.

Now Kurt Liffman and his colleagues at the University of Melbourne are trying to devise a computer model for a particulate pile, which mathematicians call a "sand pile". However, they are finding it difficult to duplicate the precise pattern of forces discovered by the two Czech experimenters.

Liffman's computer model starts with a two-dimensional pile of hundreds or even thousands of particles distributed in the pattern of an equilateral triangle. The pile is then allowed to settle. This involves solving the frictional and motion equations for each particle in an iterative way until the pile is stable.



Sand pile puzzle: why is the greatest pressure not at the centre of the base?

A dynamical computer model for a sand pile is hard to set up because the equations describing the dynamic forces between as few as three objects cannot be solved exactly. So-called N-body problems, where N is 3 or more, have long resided in the "too hard" basket of classical mechanics, and require some clever approximating techniques.

Liffman and his colleagues find that the shear forces predicted by their model agree with experiment, but they cannot yet explain the central dip in downward pressure. They speculate that extending the model from two to three dimensions may help, but there remains the possibility that a mechanism that they have not forseen is operating.

It is important to understand such piles of particles because they are widespread in industry. A pile of ore or fertiliser may cake or fragment under its own weight, and it may be important to be able to predict the conditions under which this will happen. The mathematicians have submitted their work to the journal Powder Technology.

The Melbourne group's sand pile is part of a broader project using similar techniques to explore the force patterns between particles in suspension. Their results may ultimately be of interest to the mineral extraction industry. □

The main objective of this thesis is to assess the viability of using EIT to measure pressure in a practical medical environment. However, will it replicate the Czech scientists' perplexing pile of apples pressure distribution?

2. Pressure-Sensitive Materials

The proposed EIT system requires a material which exhibits known pressure-resistivity characteristics. This chapter presents two such practical materials, namely conductive elastomers and carbon fibres. The characteristics of each are analysed and the chapter concludes with the most suitable mat construction for use in an EIT pressure scanner.

2.1. Theoretical Pressure-Sensitive Materials

Excluding the effects of shear, an interface pressure is directly attributable to the transverse force exhibited onto the subject and is defined by the following equation

$$p = \frac{F}{A} \tag{2.1}$$

where p is the interface pressure, A is the area over which the transverse force, F is uniformly applied. Knight[1] showed that if this force is applied to a thin conductive material, there will be a theoretical deformation in thickness of the material, which in turn will cause an increase in the material's resistivity. In theory, any conductive material will have the necessary pressure-resistivity characteristics to implement an Electrical Impedance Tomography based pressure scanner. However, a search for a conductive material that will deform elastically and with measurable resistivity changes due to a change in the material's thickness only, has proved to be fruitless. Many conductive materials are metals, which do not deform easily with the types of loads expected in the targeted applications. The requirement for a flexible material also rules out the possibility of using such materials.

Piezoelectric materials have been used in pressure measurement. However, piezoelectric materials produce a signal (voltage) only under dynamic situations. As static pressures need to be measured, they are not suitable. The following sections examine how various other conductive materials change resistivity when a pressure is applied. The two types considered are piezo-resistive materials; materials based on conductive elastomers and carbon fibres.

2.2. Conductive Elastomers

Elastomers are polymers or rubbers that when subjected to a force, within their elastic limit, will return to their original shape after the force is removed. The rubber is made

conductive by mixing in either carbon or metal particles at the time of manufacture. These rubbers are also pressure sensitive. The mechanism by which these materials increase their conductivity under a load, is the joining together of chains of the carbon or metal particles. An overview of these pressure sensitive rubbers, including their structure, historical development, manufacture and applications are given by Sakamoto[15] and Nagata[16]. Mokshagundam[17] provides an overview of various single pressure sensor constructions and discusses the advantages and disadvantages of conductive elastomers, which are summarised in the following paragraph.

The advantageous characteristics of conductive elastomers are their resilience or elasticity, durability and adequately wide operating temperature range. Silicone rubbers are not noticeably susceptible to ageing, oxidation and other weathering influences. However, there are a larger number of disadvantages in the important electrical characteristics. These include different values of resistance depending on whether the material is being loaded or unloaded. This hysteresis can be considerably reduced if the rubber is compounded properly. A closely related problem is the fluctuation of the no-load resistance after loading, also known as instability. Creep, the gradual change in resistivity under a constant load, is a further drawback. Another common difficulty is making reliable connections to the elastomer.

The possibility of using conductive elastomers was undertaken. An industrial company, Leyland & Birmingham Rubber Company, supplied a sample of epichlorahydrin doped with a superconductive carbon black. To overcome any connection problems, the sample was supplied with a thin layer of copper embedded into both top and bottom layers. The sample was loaded with varying weights, then unloaded. This cycle was repeated half an hour later, the results of which are shown in Figure 2.1. The sample exhibits hysteresis and large non-repeatability. A natural rubber would reduce the hysteresis and the company felt they could improve on the repeatability, through different mixing and processing methods. Other methods, such as mechanical preconditioning by cyclic loading[18], may also make improvements. However, it was felt that using carbon fibres would be a less expensive and more successful development route to take. The results from the sample confirmed many of the disadvantages given above.

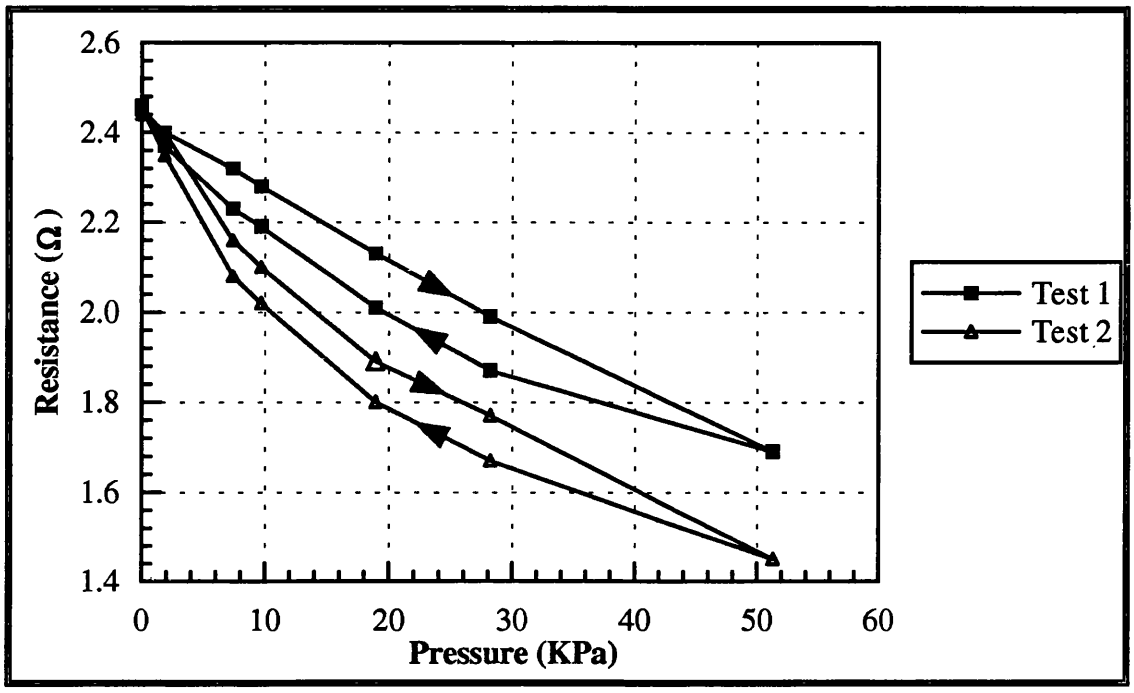


Figure 2.1 Loading and unloading tests on conductive elastomer sample

The Gates Rubber Company Limited supplied a ½mm thick sheet of silicone loaded with carbon black, which they called piezo-resistive elastomer. The sample came with resistance-pressure data curves similar to the Leyland & Birmingham sample when measured from top to bottom of the sample. Attempts were made to measure the resistance across the sheet. Copper clamped to two ends of the sheet acted as electrodes. However, no sensible results could be obtained. Resistance changes across the sheet were noticed whenever the material was stretched.

Basarab-Horwath[19] recently published a paper also suggesting the use of EIT to measure interface pressures. A conductive elastomer was proposed for use as the sensor element. The main problem they encountered was the formation of good reliable electrical connections to the elastomer. Their solution involved screwing flat ended brass screws down onto the elastomer's surface to ensure a firm connection. They found that voltage readings at the electrodes did change with varying loads and that they were repeatable. Their results were initial results into using EIT for pressure measurement and are encouraging. However, the following section describes a mat made from carbon fibres that would be more suitable for an EIT pressure measurement system.

2.3. Carbon Fibre Felt

Carbon fibres are fine strands, cylindrical in cross section and are made by carbonising certain organic fibres. Carbon fibres are commonly used for their high strength to weight ratio in aircraft, boat, tennis racquet etc. construction. They are also electrically conductive and a single carbon fibre's resistivity can vary from 10^{6} to 10^{-3} Ω.cm[20]. Carbon fibres are produced by oxidising and subsequently graphitising acrylic fibre. This basically involves heating in air before heating and stretching the oxidised fibres in a nitrogen furnace.

The fibres can be manufactured into thread, felt, bundles or woven cloth. Felt displays useful pressure-resistivity characteristics and is almost exclusively used in carbon fibre pressure sensors[17]. Katahira[20] showed that the resistance of a piece of carbon felt versus applied force is logarithmic, when the resistance is measured in the same direction as the applied force. Pruski[21] shows that resistance also changes with applied pressure, when the resistance is measured across the felt, i.e. orthogonal to the direction of force. It is this type of pressure-resistivity characteristic that fulfils the requirements for EIT pressure measurement. Carbon felt decreases its resistivity as pressure on it increases and the reasons are dependent on the degree of applied pressure[22]. The structure of carbon felt, described below, explains this behaviour.

Carbon felt simply consists of many fibre strands of say 2mm in length and about 10μm in diameter, randomly intertwined or compacted into a matrix. This results in the same sort of structure as cotton wool and the felt can be manufactured as thick as required. In the initial stages of increasing load, a decrease in resistance is attributed to an increase in the number of current paths due to an increasing number of contacting fibres. At higher pressures there is an increase in the contact area between fibres. Carbon fibre felt has successfully measured force using individual sensors in gait analysis[23]. Robot tactile sensing by multiplexing strips of carbon fibre felt has been achieved by Robertson[24] and Pruski[25].

2.3.1. Characteristics

Carbon felt has a very high dynamic range of resistance, at least 4 decades[22]. The range of pressures over which it is sensitive is also very wide. The good elasticity and great strength of the fibres enable very high pressures to be measured before any mechanical breakdown occurs and also enable a smooth change in resistivity[22]. Larcombe[22] also

found the material repeatable within 1% even after the felt had been subjected to a 200MPa load. Katahira[20] also showed good repeatability and stability.

Katahira[20] looked at the voltage gain/frequency characteristics of his carbon felt sensor. The frequency response is a low order decay with a break frequency of 30Hz. The frequency at which current patterns can be applied to an EIT system will be limited because of the decreased response at high frequencies. This in turn limits the frequency at which entire pressure scans are taken. However, not many of the envisaged medical applications mentioned require scanning at high frequencies. Mokshagundam[17] examined conductive elastomers and carbon fibres for use as pressure sensors and states that carbon fibres have a low hysteresis, lower than conductive elastomers.

The following section discusses why carbon felt is the chosen material for the proposed EIT pressure scanner.

2.4. The Proposed Mat Construction

A mat constructed from carbon fibre felt which is encapsulated in a gel or silicon rubber is the most appropriate manner to measure pressure in an EIT system. Figure 2.2 illustrates the mat construction. The mat would not be rigid and would thus conform to the shape of the interface at which the pressure is being measured. A flexible as well as thin mat introduces the minimal amount of distortion in the true pressure profile. Many available pressure sensors are rigid and are thus unable to measure pressure in the targeted applications.

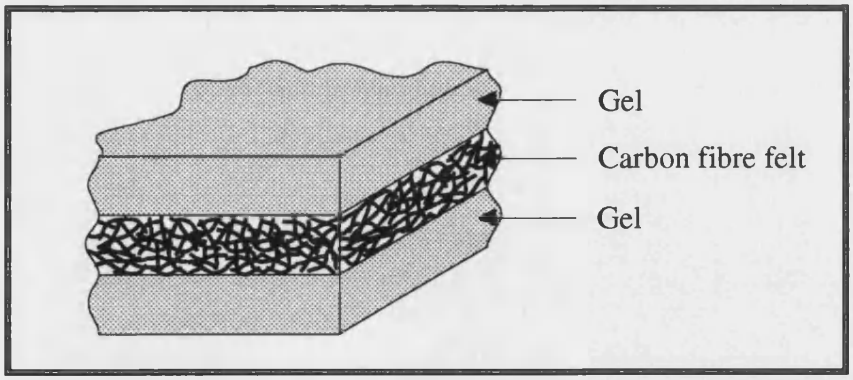


Figure 2.2 Proposed mat construction showing the layer of carbon fibre felt encapsulated by thin layers of gel.

A mat constructed from carbon fibres would be inexpensive. Felt is fairly cheap (~£40m^{-2}) and encapsulating the fibres within a rubber is a simple task. By comparison, conductive elastomers cost around £1200m^{-2}. Accurate placing of the connections to the mat is not a problem and can also be done inexpensively. Larcombe[22] established the best electrical connection to the felt by using a light hardening adhesive; the same can be applied for the point electrodes required in the EIT system. The natural strength of carbon fibres ensures that a sensor mat made from them will be rugged. However, abrasive decay may occur if the fibres are not encapsulated in a rubber. This rubber must not react with the fibres as this would cause a change in their electrical characteristics[17].

Under very low loads, there is a certain amount of noise generated from microscopic movements between the fibres. However, with a 10Pa load, Larcombe[22] almost nullified it. If the carbon felt is encapsulated within a gel, the weight of the gel will have the same effect. Two other reasons for using an encapsulating gel are to hold the fibres in place and to render it impervious to bodily fluids, ensuring easy cleaning of the mat.

Different sized mats may have different electrode spacings. The wide range of available fibre resistivities ensure an appropriate fibre resistivity can be selected. Optimal choice in resistance between mat electrodes will minimise implementational difficulties in the driving and sensing electronics.

In summary, carbon fibres have been chosen for the EIT system as they have the desired electrical properties to measure pressure. Additionally, they have low hysteresis, are thermally stable and give repeatable results. The fibre's mechanical properties give rise to a strong durable mat. The proposed mat construction will overcome the carbon fibre's limitations of noise and abrasive decay.

3. Introduction to Electrical Impedance Tomography

Electrical Impedance Tomography is a technique which is able to construct an image of the distribution of electrical impedance, within an electrically conducting object, from knowledge of voltage and current conditions at various points on the surface or periphery of the object. This chapter gives an overview of EIT, covering some of the more popular algorithms in use. The algorithm which reconstructs the pressure distribution is introduced.

3.1. Background to Electrical Impedance Tomography

Electrical Impedance Tomography is a new field of imaging in which an increasing amount of research has been undertaken over the last decade. Impedance imaging has three primary applications, namely in the medical, geophysics and industrial fields. In industrial applications, Process Tomography uses EIT to image conducting fluids within vats or flowing within pipes[26]. Typical geophysics applications are for obtaining core sample data, geophysical probing for detection of minerals at the earth's surface, and in borehole scanning. In the medical field, interest is primarily in imaging body organs, bones and other tissues by placing electrodes on the surface of the skin. This is possible as the various body tissues have different characteristic impedances or more specifically resistivities[27]. An image is obtained by taking one or more two-dimensional slices through a patient. EIT is one member of a whole class of imaging techniques in which the common objective is to non-invasively determine certain physical parameters within a region of interest. Some of these techniques include X-ray Computed Tomography, Nuclear Magnetic Resonance Imaging, B-scan Ultrasound and Positron Emission Tomography. A comprehensive comparative review of such techniques has been published by Bates[28].

Electric Impedance Imaging began as an offshoot from X-ray Computed Tomography to overcome its disadvantages of expense and biological hazard to the patient being imaged. The first attempt[29] based on X-ray Computed Tomography is flawed in its theoretical assumption that electric currents flow in straight lines, in the same manner that beams of X-rays[30] do. Since this initial attempt at impedance imaging, a whole range of image reconstruction techniques have been developed specifically for electrical current behaviour. A recent review of these has been published by Isaacson[31]. Many of these techniques are theoretical methods and it is only recently that *in vivo* images have

been produced. A more comprehensive study of the evolvement of EIT may be obtained from Knight[1].

3.2. The Various EIT Reconstruction Algorithms

Reconstruction algorithms for conductivity imaging can be broadly divided into those that seek a full reconstruction using a Newton approach, and those that linearise the problem using a modified backprojection technique. The latter approach usually gives a less accurate, but a faster solution. The Newton routines are nonlinear methods and tend to give more accurate reconstruction, but this is on a theoretical level using computer simulations. Most of the images obtained from human data is obtained using the linearised backprojection method of Barber[32]. Conductivity images published using this method have not shown particularly good spatial resolution[27]. The method can be used for dynamic images with good contrast, good sensitivity to resistance changes, high speed and low cost[33]. However, it was for quite some time the only method to reconstruct *in vivo* images.

Recently Woo[34] published a paper using an improved Gauss-Newton reconstruction algorithm. The paper attempts to measure lung resistivity for detection and monitoring of apnoea and oedema. It is the first paper published which demonstrates the feasibility of this algorithm using a computer simulation as well as images from a human being. The backprojection method of Barber[32] has been used for the same application of imaging resistance changes during respiration by Harris[35]. However, the fundamental limitation of this system is that it cannot measure absolute resistance, only a change in resistance during the respiratory cycle[34]. The system of Woo[34] measures an absolute lung resistance, in an attempt to quantify the amount of air and water present in the lung. Despite the disappointing accuracy in lung resistivity that they measured, the authors have confidence in the method used and propose solutions for improvement. More recently, reconstructions using the Newton approach successfully monitored lung resistivity changes during the formation of pulmonary oedema[36].

There are limitations of the Newton approach for a practical implementation, as the inaccurate lung resistivities show. However, many of these are removed when using a conductive mat, as opposed to a human being, as the usual practical errors do not apply. These are discussed in the following section.

3.3. Previous Work on EIT Techniques to Measure Interface Pressure

Knight in his thesis[1] reviewed many of the EIT reconstruction techniques specifically for measuring pressure distribution on a conductive mat. He reviewed five different reconstruction algorithms for a pressure scanner. These were the reconstruction assuming a linear forward transformation of Yamashita[37], the Algebraic Reconstruction Technique of Kim[38], the Simultaneous Iterative Reconstruction technique of Yorkey[39], the Gauss-Newton method of Yorkey[40] and the backprojection method of Barber[32]. This last EIT reconstruction technique is also known as Applied Potential Tomography (APT).

The first three algorithms mentioned above gave very poor results in the computer simulations used. This, Knight concluded was due to the assumptions of these three methods, that there existed some kind of linear relationship between the changes in the resistivity profile and the voltages appearing at the peripheral nodes of the simulated profile. This nonlinearity was clearly shown by him and so these algorithms were discarded for good image reconstruction.

The latter two algorithms, which have already been mentioned, were the most hopeful. Out of these two, the Gauss-Newton method was concluded to be the most suitable for use in the proposed pressure measuring system for the following reasons:

- The Gauss-Newton method was the only one to give consistently reliable and accurate results over a wide variety of resistivity profiles, even with simulated voltage error measurement. Knight noted that the reconstruction is markedly affected by these errors. However, this thesis details refinements to the Gauss-Newton method which greatly improve its error susceptibility.

- The backprojection method is less susceptible to measurement errors than the Gauss-Newton method. One of the major causes of measurement errors is the use of two-dimensioned resistivity models in a three-dimensioned system. Three-dimensional current flow is removed when using a mat as opposed to a human subject when mapping resistivity. The modelling is thus more accurate for a resistive mat and the better error tolerance advantage of the backprojection method is reduced.

An additional note on errors is that *in vivo* errors, normally introduced by variable electrode placement on human subjects, will be very low on a fixed electrode mat. Hu[41] when looking at all the physical errors in practical EIT systems, concluded that electrode position error is one of the largest sources of measurement error. They therefore suggested using a precise layout of the electrodes in any physical phantom model. This is possible using the proposed pressure measurement system in the clinical environment during the manufacture of the mat. For medical applications, Gencer[42] even suggested immersing the object/subject to be imaged into a cylindrical container filled with water of known conductivity. Electrode placement errors are minimised as the electrodes attached to the cylindrical container are in permanently fixed positions. Ultrasonic techniques are then used to detect the object's cross-sectional shape so that the varying distance between the object and the container wall can be measured and used. This extra effort demonstrates the importance of placing the electrodes in fixed positions.

3.4. Introduction to the Reconstruction Process

The reconstruction problem is to solve a resistivity (conductivity) distribution within a medium, given known voltages and current densities. The governing equations for this problem, when the medium is inhomogeneous and anisotropic, is given by

$$\nabla \cdot \rho^{-1} \nabla V = 0 \tag{3.1}$$

where ∇ is the Poisson operator, ρ is the resistivity and V is the voltage distribution within the region being studied.

The reconstruction process is broken into two problems, namely; a forward and inverse problem. Figure 3.1 shows the definition of the forward problem; find the internal voltage and current density distributions given the resistivity distribution and certain current and voltage distributions.

The inverse problem is defined by Figure 3.2; given the boundary voltage and current density distribution, solve for the resistivity distribution.

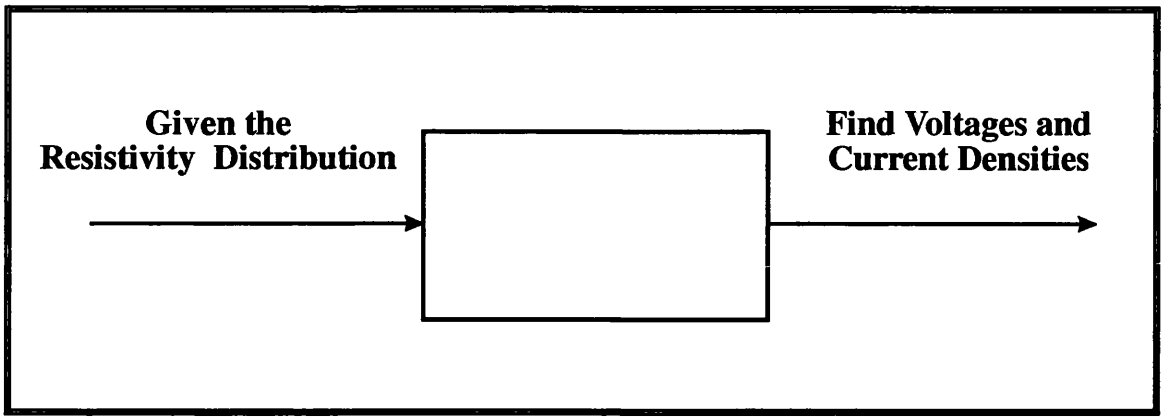


Figure 3.1 Representation of the forward problem

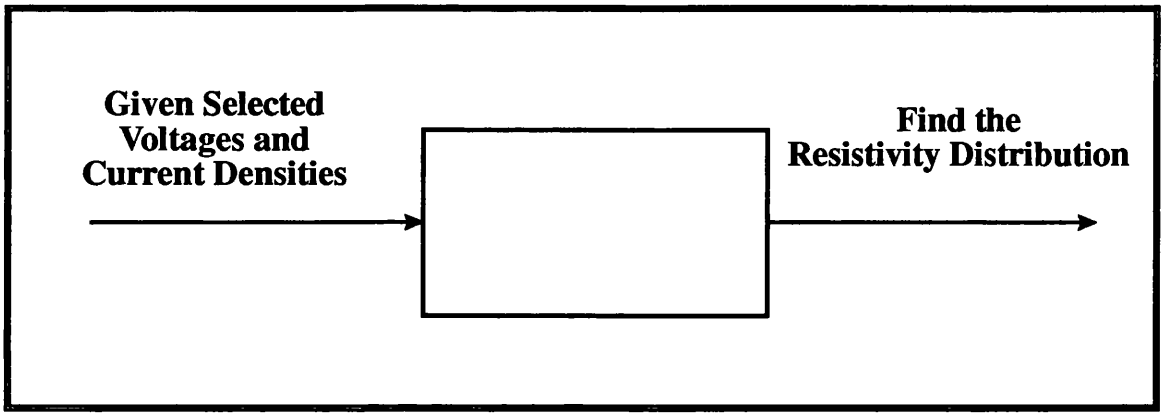


Figure 3.2 Representation of the inverse problem

To solve for the inverse problem (the defined problem), a numerical solution is employed, using iterative solutions of the forward problem. Solving for the internal resistivity distribution is a boundary value problem as only selected voltage and current density distributions are known. As many of these voltages and currents on the mat are unknown, the problem is a nonlinear inverse problem. To turn pressure measurement on a continuous resistive mat into a linear problem, wires to each node on a finite element mesh would be necessary. This might mean 441 wires for a 400 element mesh, a practically difficult situation. The EIT approach only requires a manageable 30-40 electrodes to obtain pressure reconstructions.

The governing equation (3.1), is a complicated differential equation and it is impossible to find its general analytic solution. The finite element method is a numerical technique to solve equation (3.1), given the appropriate boundary conditions. The following

chapter describes the manner in which the finite element method is implemented to solve the forward problem. Optimisation techniques include the least squares method, under which the Gauss-Newton method can be classified. Chapters 5 and 6 study optimisation techniques and expand on the basic Gauss-Newton method so that the inverse problem can be solved in a practical environment. Chapter 7 examines electrode positions and is the final area to be studied in this thesis.

4. The Forward Problem

In Electrical Impedance Tomography there is both a forward problem and an iterative inverse problem to solve. The forward problem is based on the Finite Element Method (FEM) and this chapter is concerned with the solution of the finite element equations. The finite element model is a mesh based on a square domain comprising square bilinear elements. Justification of this model and the implementation of the automatic mesh generator is discussed. The fast solution of the finite element equations is important as it reduces the total time required to solve for the pressure distribution. Both sparse matrix methods and sparse vector methods reduce the finite element equation solution time. The most appropriate sparse techniques are studied and implemented.

4.1. FEM Theory

The finite element method, rather than the finite difference method (FDM), is chosen to solve the forward problem from the arbitrary resistivity distribution. The FEM gives a piecewise approximation to the governing equation (4.1), whereas the FDM gives a pointwise approximation. Therefore, most EIT reconstruction algorithms use the FEM to solve the forward problem[43]. The FEM is also superior to the FDM when modelling arbitrary shapes and this is another reason why the FEM has been almost universally adopted by EIT researchers.

In the FEM, the calculus problem, defined in equation (3.1), is solved by changing it into the algebraic linear system of equations:

$$\mathbf{Yv} = \mathbf{c} \tag{4.1}$$

where $\mathbf{Y}$ is the $r \times r$ admittance matrix having units of conductance, $\mathbf{v}$ is the measured volts vector, $\mathbf{c}$ is the vector containing the applied currents and r is the number of points for which $\mathbf{v}$ and $\mathbf{c}$ are solved.

The FEM divides the continuum region or domain into a finite number of geometrical shapes or elements. In the implementation used the elements are square and the domain is the resistive mat. Each element is assumed to have an isotropic and homogeneous resistivity. This discretisation of the continuous problem expresses the unknown field variables in terms of certain interpolation functions within each element. An example of one of the meshes used with 25 elements is shown in Figure 4.1.

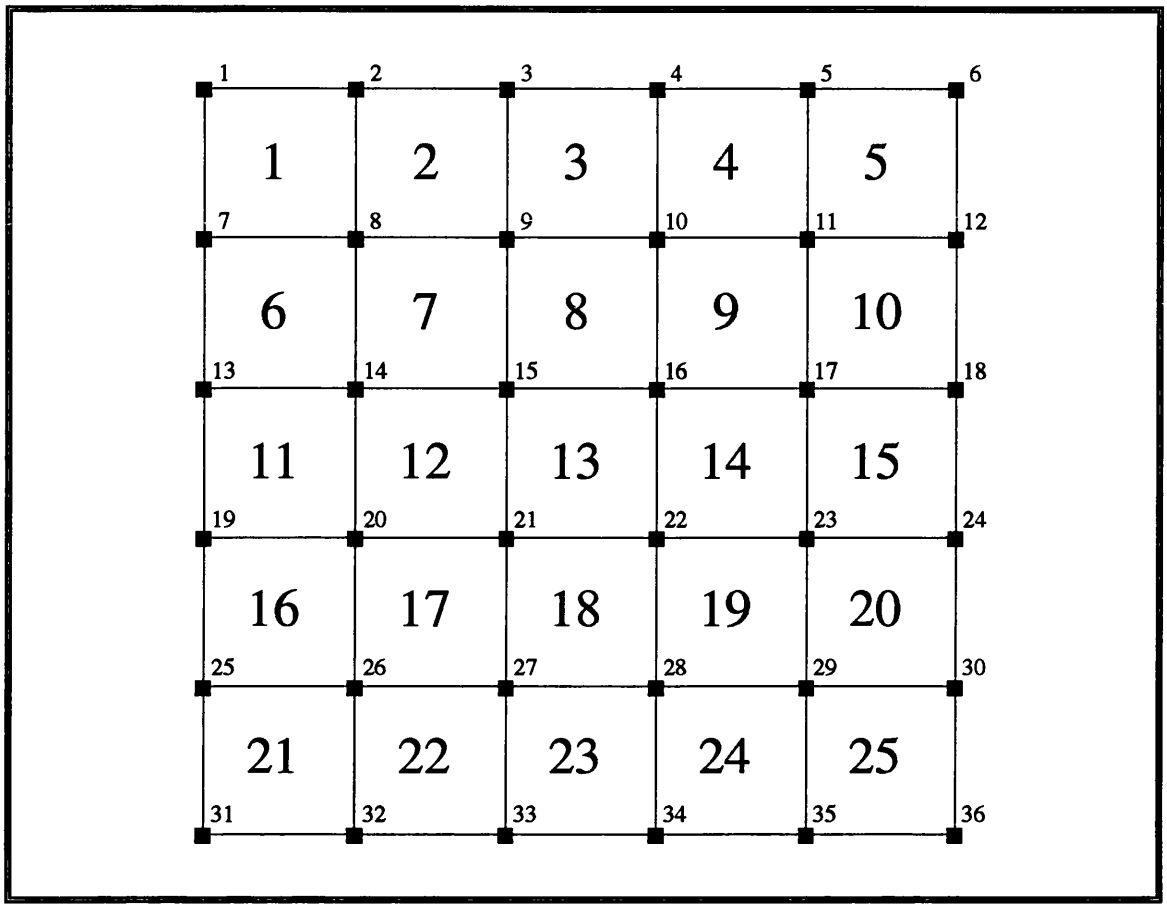


Figure 4.1 A FEM mesh of 25 elements with the natural (row by row) node and element numbering

A single isotropic homogenous element with node numbering is shown in Figure 4.2(a). A resistor network could be used to represent this element. Two such possibilities are shown in Figure 4.2(b) and (c).

When Knight[1] implemented the Gauss-Newton method to obtain his practical results, he chose to use the resistor network in Figure 4.2(b) to model the resistive mat. He used the physical procedures utilised in network analysis to obtain the associated element equations needed by the FEM. Huebner[44] called this the direct approach for determining element properties. A better finite element model for the resistive mat can be used. A mathematical approach to find this model could be to use the variational principle to derive the element equations. Another mathematical approach is the method of weighted residuals with Galerkin's criterion. Identical element equations are obtained for Poisson's

equation when using either approach[44]. The variational approach is now shown in some detail to produce a more accurate finite element model for an element. The model derived, is in fact an analogy to the resistor network in Figure 4.2(c) and is known as the bilinear element.

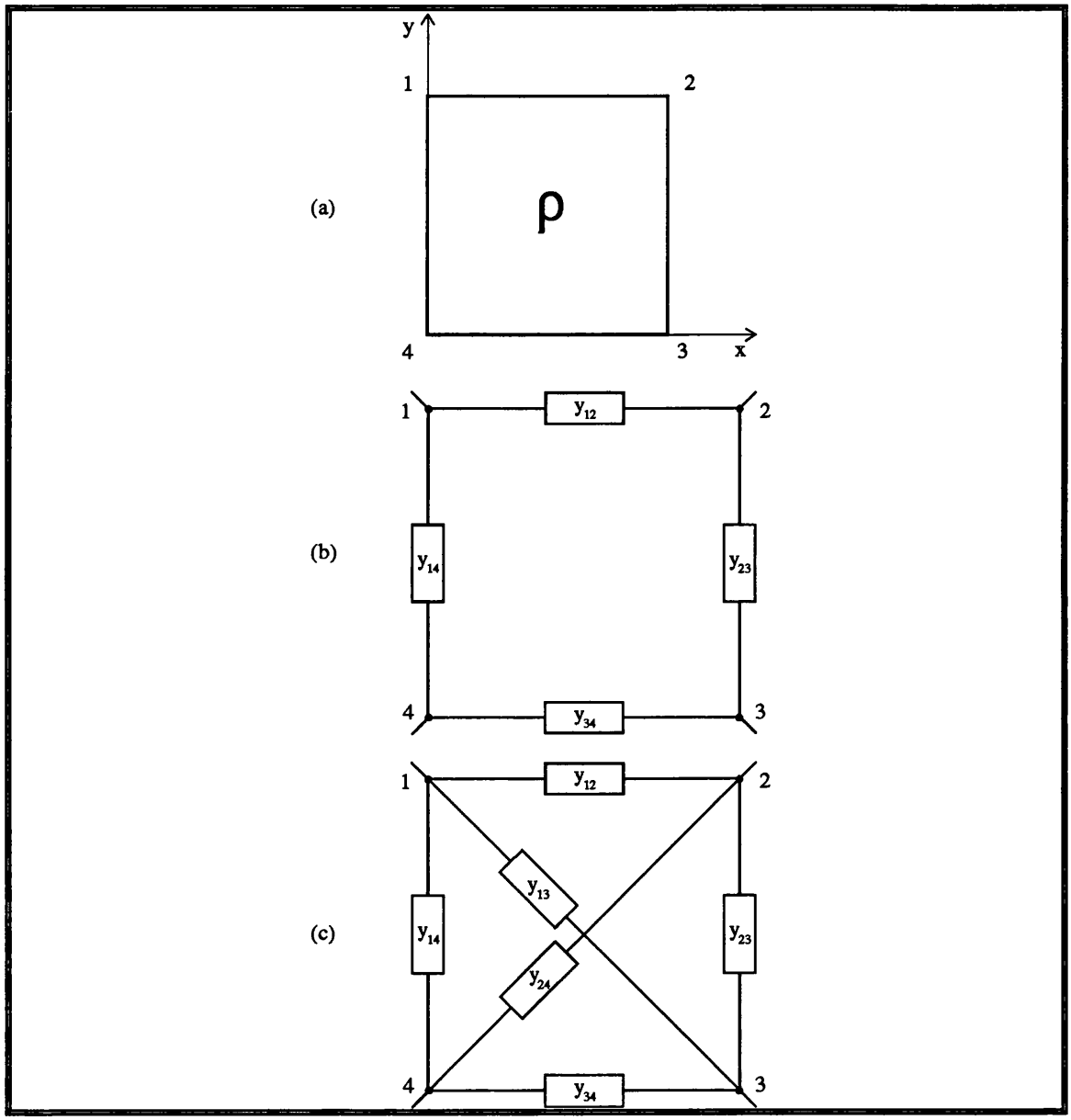


Figure 4.2 (a) Single element of resistivity ρ, with local numbering
(b) A resistor network used to represent the element in (a)
(c) A better resistor network to represent the element in (a)

4.1.1. The Mathematical Interpretation of the FEM

In a physical interpretation of the FEM, the elements are thought to be individual segments or parts of the actual system. In the case of a resistive mat, the domain is imagined to be comprised of many linked resistor networks. The mathematical interpretation of the FEM requires a generalised definition of the elements and to approach the problem in less physical terms. Instead of viewing the element as a physical part of the system, the element is viewed as a part of the solution domain where the phenomena of interest are occurring. Once a mesh for the solution domain has been decided, the behaviour of the unknown field variable (voltage), over each element is approximated by continuous functions expressed in terms of the field variable at nodes of the element. The functions defined over each finite element are called the interpolation functions. The collection of interpolation functions for the whole solution domain provides a piecewise approximation to the field variable.

Appendix A4 details the Finite Element Method for obtaining the voltages at each node on the mesh. Variational principles are used to derive the element equations. Once the equations for each element are known, they must be assembled to represent the entire domain. In assembling the element equations to form the master stiffness matrix, the boundary conditions are implemented. The boundary is taken to be any node that is available for voltage measurement or current injection, i.e. the electrodes. This includes those on the physical boundary or edge of the mat, as well as those used in the middle of the mat.

The bilinear element stiffness matrix is evaluated in Appendix A4 and shows that it has units of Siemens, the same as an electrical network admittance matrix[45]. Yorkey[46] showed that using such elements to assemble the master admittance matrix solves the FEM using an appropriate resistor network and that this matrix satisfies Kirchoff's Laws[47], even if the network is not physically realisable. The bilinear element stiffness matrix does in fact represent the network shown in Figure 4.2(c). The direct approach demonstrated by Knight[1] can also be used to derive the element stiffness matrix for this particular resistor network. The element equations have been derived from variational principles here and the resistor network that can be associated with these equations is a better representation of the resistive mat than the network in Figure 4.2(b).

4.1.2. Discretisation of the Domain

When discretising the continuum by dividing it into a finite number of elements, the main aim is to achieve a good representation of this continuum. The type of element to be used is of importance. There are no set rules for achieving this aim, but some general guidelines are available from experience in finite element analysis. The best guideline is from previous work done in the same field of interest, as construction of suitable elements and meshes requires skill and ingenuity stemming from much experience[44].

4.1.2.1. Element Shape

There are many competing types of elements that are used to model domains of interest. It is not clear whether it is more efficient to use triangles or quadrilaterals when subdividing the domain[48]. The most often used element is the simplest two-dimensional element, the three-node triangle. The reason for this is that an assemblage of triangles can always represent a two-dimensional domain of any shape. Much of the latest research in EIT uses triangular elements to model the curves in the human body. In doing so, the geometric error in the domain model is minimised. Hua[49] used a fairly complicated mesh of triangles, rectangles and quadrilaterals in his EIT model. This mixture of element shapes is quite normal in finite element modelling. Quadrilaterals and especially rectangles are enjoyed for use in the model's interior as fewer are needed to fill a region and they permit simple elements of high degree with the added advantage of a 'twisting' term[48]. The pressure sensitive mat may be cut to any desirable shape and it would thus seem pertinent to use a rectangular or square shape so that quadrilaterals might be used to model it, both for the interior and at the mat edge. With accurate cutting of the mat, the geometric modelling error will be almost zero.

The continuum should be divided so that the elements have a well-proportioned shape, i.e. avoiding long, narrow elements. For a given element in a given location, the optimum size and shape depends on the local gradient of the voltage[44], but this is unknown *a priori*. A conservative procedure would be to use the square elements which are well-proportioned.

4.1.2.2. Selection of Interpolation Functions

Polynomials are used as interpolation functions as they are easy to differentiate and integrate. Many different types of interpolation functions exist and Strang[48] examines the different rectangular elements that are in popular use.

The simplest square element, shown in Figure 4.3(a), corresponds to a bilinear interpolation function. Improvements in modelling accuracy can be obtained if higher order interpolation functions are used. The next highest order interpolation functions are the 8 and 9 noded elements as shown in Figure 4.3(b) and (c). In EIT, the maximum number of elements is restricted by the number of electrodes being used, see the following section. The accuracy of the Jacobian matrix and the voltages which the FEM model produces is thus limited unless higher order interpolation functions are used. However, the low order bilinear elements are simpler to implement and require a smaller element stiffness matrix, which enables faster solution of the finite element equations and calculation of the Jacobian matrix. Errors in the finite element model may lead to the Gauss-Newton method converging to local minima. Any trade-off between accuracy and time to solution depends on how successful the interpolation functions model the domain. This is not known and indeed, a fine mesh comprising of many simple bilinear elements approach, may or may not provide a small enough modelling error. This depends on other errors, such as those in the electrical system or accuracy in electrode placement. Paulson[50] and Strang[48] show mathematically the rate of convergence in the modelling error for different order interpolation functions, but precise evaluation can only be achieved once the other errors are known. In other words, when the voltages measured from a real pressure-sensitive mat can be measured and compared to the calculated values. Most other EIT researchers have used first order interpolation functions (linear triangles and/or bilinear quadrilaterals) successfully.

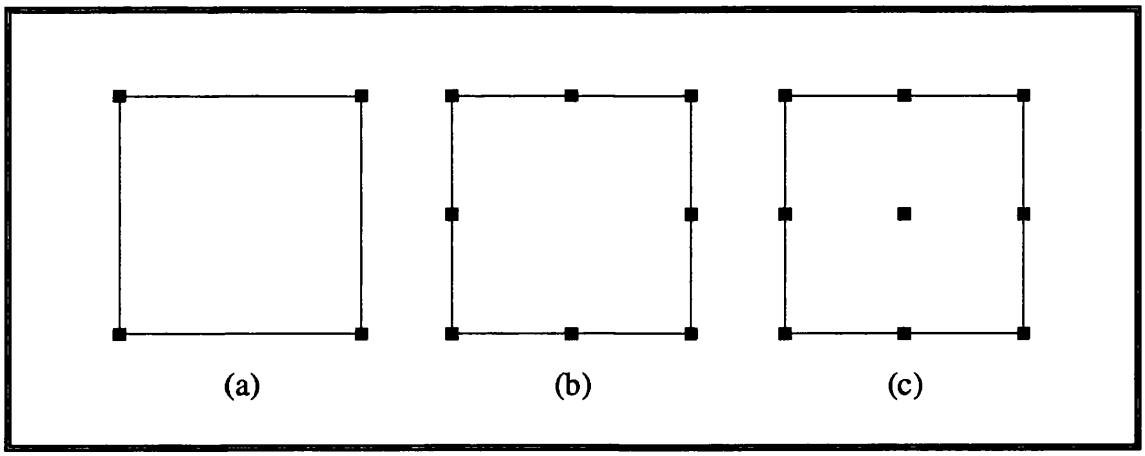


Figure 4.3 (a) The simple 4 noded bilinear element
(b) The restricted biquadratic square element
(c) The biquadratic square element

4.1.2.3. Mesh Coarseness

Greater spatial resolution is obtained near the surface than at the centre of a body in normal EIT applications. In this case, a finer mesh ought to be used at the surface and a course mesh at the centre, such as the one used by Paulson[50]. This is because the spatial resolution, defined as the smallest region of a medium into which the conductivity can be isolated[51], is much lower in the centre of the body than at the edge. In the design of a mesh for pressure measurement, the spatial resolution is more uniform across the mat due to the electrodes being positioned throughout it. The mesh will give a good representation of the real-life situation by using equi-sized elements across the domain.

The finer the mesh, the higher the resolution of the pressure distribution and the more superior the image. As the number of elements increases and the size of the elements in the mesh approaches zero, the finite element model approximation approaches that of the true distribution. In the real world, very high resolutions are unwarranted and impose unnecessary and difficult implementation problems. Finer meshes are computationally more expensive when using the Gauss-Newton algorithm. A trade off between resolution and pressure measurement intervals must be made. This is application dependent and for a hospital bed, a resolution of 5 to 10 mm square would suffice.

The maximum number of elements is also limited by the number of independent measurements. Where m is the number of elements and e is the number of electrodes, this maximum is[32]

$$\max(m) = \frac{e(e-1)}{2} \tag{4.2}$$

Since this is equivalent to the maximum number of equations in the Gauss-Newton method, the number of maximum resistivity values is also limited by this number. So, the spatial resolution of the resistivity image is primarily limited by the number of electrodes. In a practical environment, the number of electrodes must be limited to reduce the number of undesirable electrode wires trailing from the mat. For a scheme of 32 electrodes, the maximum possible number of elements is then 496. Even with this limitation, the computing time can be lengthy.

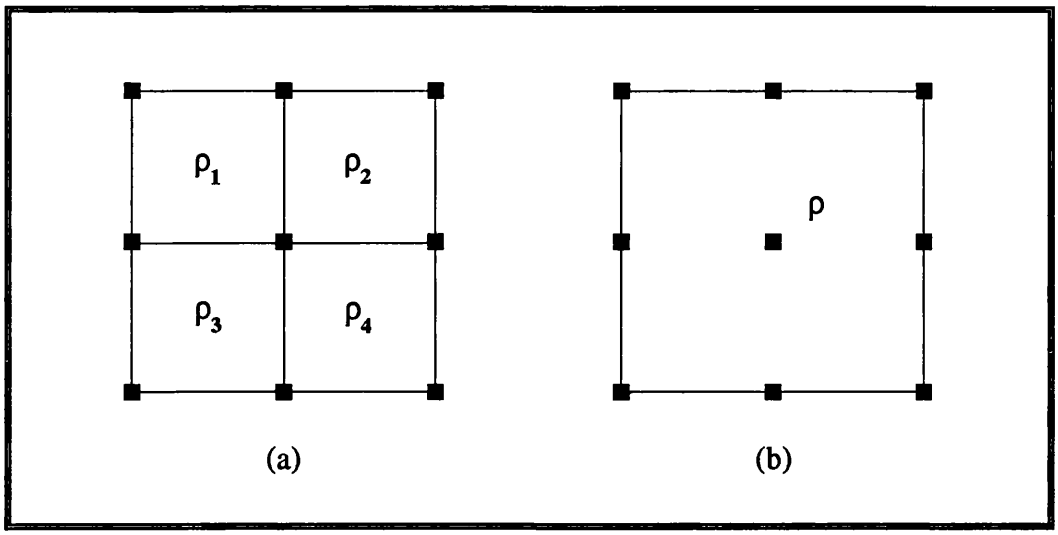


Figure 4.4 (a) 4 bilinear elements
(b) 1 biquadratic element

The mesh coarseness chosen need not be this maximum and depends on computing time limits and the modelling error introduced, that is, the accuracy of the Jacobian calculation and the electrode voltage approximation. Two approaches could be taken. The first involves creating a finer mesh by dividing elements into four smaller bilinear elements as shown in Figure 4.4(a). The second uses the higher order biquadratic interpolation function for an element, as shown in Figure 4.4(b). Both have the same number of nodes and produce the same size master stiffness matrix, leading to similar forward modelling computation times. However, when the entire mesh is examined, 4 bilinear elements give a higher spatial resolution of resistivity than 1 biquadratic element. Both reduce the modelling errors, but the latter has the disadvantage of a lower resolution in the resistivity solution. The former case requires the solution of 4 times as many resistivity equations and so the G-N update takes longer, but the advantage is a higher resolution. The division of each element into 4 further bilinear elements must take into account the maximum number of equations defined by equation (4.2).

The number of elements in the forward modelling is not restricted by equation (4.2) and so more elements could be used for this part alone. In other words, a fine mesh could be used for the forward modelling and a coarse mesh for the G-N resistivity calculation (inverse problem). Hua[43], in his circular mesh, used more elements than that allowed by the maximum for the forward problem. In the inverse problem he constrained elements in the outermost area of the mesh, where high resolution is not necessary, by assigning

the same resistivity value to groups of elements. In this way, the number of elements in the mesh is, in effect, less than the maximum allowed.

4.2. FEM Implementation

This sub-section starts by examining the automatic generation of different sized finite element meshes in modelling the mat's resistive distribution. The validity of these models is established. The bulk of the sub-section covers the practical methods of solving the finite element equations produced by the FEM model. The steps to solution are broken into different stages and are looked at individually. Of prime consideration is the reduction in computer processing time in arriving at the equation's solution.

4.2.1. Mesh Generator

The chosen mesh comprises of square elements to model a square domain. A mesh generator is needed for flexibility so that different sized finite element meshes may be used to model the resistive mat. The mat can be discretised into a range of meshes, from a coarse to a very fine mesh. The implementation of different finite element meshes is laborious and is best done by an automatic mesh generator. A simple mesh generator was programmed to produce the required data files needed for the finite element model. A uniform mesh composed of equi-sized square elements for a square shaped domain is produced. The mesh generator input is simply the desired number of elements down the side of the square domain.

The mesh generator produces two data files. The first holds the element stiffness matrix, as shown in equation (A4.15) and the node topology. The node topology is a table to transform the local node numbering to the global numbering for the assembly of the master stiffness matrix **Y** from the element stiffness matrix. This table is sometimes called the connectivity table for the mesh. All the nodes and elements in the mesh are numbered from left to right, working down the mesh from top to bottom and is known as the natural ordering. A typical mesh with node and element numbering is shown in Figure 4.1. The second data file produced is the initial estimate for the resistivity distribution of the mat. This is required by the reconstruction process. Test distributions can also be produced through simple modifications of this file.

A third data file contains the experimental set-up so that different current pattern configurations (called projections) may be chosen. This is entered by hand unless the computed diagonal or optimal currents of chapter 7 are used. The file contains

information on the number of projections, the electrode node numbers, the current source nodes and the grounded node numbers. This important file is checked by the mesh generator to ensure that node numbers are valid and that the current nodes are actually electrode nodes.

4.2.1.1. Mesh Validity

The finite element mesh and equations must satisfy the completeness and continuity conditions defined by Burnett[52]. If these two conditions are met, then a sequence of approximate solutions, corresponding to a sequence of successively refined meshes, will converge to the exact solution as the element size is reduced to zero[52]. The finite element mesh chosen is valid, and has been correctly implemented, as this convergence at node voltages was observed when the mesh was subsequently refined from 25 to 100 then 400 elements.

4.2.2. Solving the FEM equations

This section considers the practical methods of solving the finite element equations. The aim of this section is to find the fastest means of solving them. The theory presented here begins with a general matrix solution and leads on to reducing the solution times through the use of sparse methods. Obtaining and solving the sparse finite element equations is broken into six different stages. The two most important stages with regard to solution time are given most attention. This is firstly, the numbering of the nodes in the finite element mesh so that sparse matrix methods may best be taken advantage of and secondly, the use of sparse vector methods. Practical implementation of the sparse vector and sparse matrix methods are considered and finally the different equation solution times are presented in order to choose the fastest solution method.

4.2.2.1. General Matrix Solution

The finite element method produces the algebraic linear set of equations

$$\mathbf{Yv} = \mathbf{c} \tag{4.3}$$

where $\mathbf{Y}$ is an $r \times r$ sparse symmetric positive definite coefficient matrix and is also known as the master stiffness matrix. The right hand side vector, $\mathbf{c}$ and the solution vector, $\mathbf{v}$ are of length r where r is the number of nodes at which the voltages ($\mathbf{v}$) are to be solved. This equation could be solved by inverting $\mathbf{Y}$; $\mathbf{v}=\mathbf{Y}^{-1}\mathbf{c}$. This form is not used as it is slow and $\mathbf{Y}^{-1}$ contains mostly non-zero values and is known then as a dense

matrix. This occurs even when **Y** has a low percentage of non-zero entries, that is, when it is a sparse matrix. The well known technique of symmetric Gaussian elimination or Cholesky factorisation is used to solve for **v**. This involves decomposing or factoring **Y** into the form $\mathbf{LL}^T$, where **L** is lower triangular, i.e. all the entries in **L** above the diagonal are zero. Equation (4.3) becomes

$$\mathbf{LL}^T\mathbf{v} = \mathbf{c} \tag{4.4}$$

The triangular systems

$$\mathbf{Lx} = \mathbf{c} \tag{4.5}$$

then

$$\mathbf{L}^T\mathbf{v} = \mathbf{x} \tag{4.6}$$

are solved. Solving equation (4.5) is known as forward substitution and solving equation (4.6) as backward substitution.

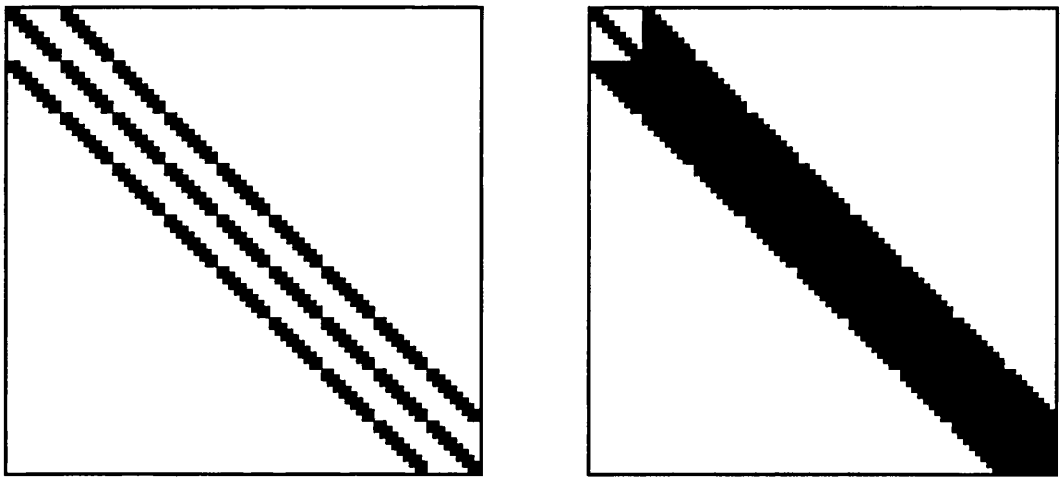
The above three equations can be solved as dense matrices, that is, not taking advantage of any special structure in **Y**. As **Y** is a symmetric sparse matrix, special sparse matrix techniques can be utilised. Sparse matrix techniques are now a well studied area of research and the various techniques in use are given a good overview by Duff[53] and George[54]. Details of the Cholesky decomposition may also be found in these books.

4.2.2.2. Sparse Equation Solution

The **Y** matrix structure results from the node ordering of the FEM mesh. The resulting **Y** matrix structure from a natural ordering on an 8×8 or 64 element grid is shown in Figure 4.5. In this and other similar figures, a non-zero value is represented by a black square and the white space represents zero values. The factorised matrix, **L** has added non-zero entries where there were zero entries in the original **Y** matrix and so has a slightly modified structure from **Y**. These added entries are known as fill-in and can be seen in the factorised matrix $\mathbf{L}\backslash\mathbf{L}^T$, (that is, $\mathbf{L}+\mathbf{L}^T$) in this figure. In this particular example fill-in of the entire area between the diagonal and the **Y** matrix non-zeros has occurred. This does not always happen and judicious choice of the node numbering can reduce the percentage fill-in quite considerably. Renumbering nodes from the original ordering is equivalent to interchanging rows and columns of **Y** through the use of a permutation matrix, **P**. The permutation matrix is simply the identity matrix with permutated rows and **Pv** is simply a re-ordering of the vector **v**. Equation (4.3) can be re-written as

$$(\mathbf{PYP}^{\mathrm{T}})\mathbf{Pv} = \mathbf{Pc} \tag{4.7}$$

Sparse techniques avoid operating on zero matrix entries and some approaches attempt to reduce the fill-in; the amount of fill-in has obvious consequences on the number of floating point operations required to solve a set of equations. This leads to the main aim of using sparse techniques; to reduce the amount of storage space used and the time spent in solving a set of linear equations. In EIT, the finite element equations and the similarly solved inverse of **Y** must be solved many times over to obtain the mat pressure distribution. The prime consideration in this repetitive process is to reduce the time taken to solve one right hand side of a linear set of equations. The reduction in computer storage space is an added bonus even though modern computers have copious amounts of RAM.



81×81 **Y** Matrix
Non-zeros: 625 Sparsity: 9.5%

81×81 $\mathbf{L}\backslash\mathbf{L}^{\mathrm{T}}$ Matrix
Non-zeros: 1521 Fill-in: 896 (13.7%)

Figure 4.5 Resulting matrix structures using the natural ordering on a 64 element mesh

The six steps to solution of the finite element equations when using sparse techniques are shown in Figure 4.6.

Note that for each repeated solution of the right hand side vector, neither the master stiffness matrix nor the factorised matrix changes. This leads to fast solution times for

different right hand side vectors. However, each Gauss-Newton (G-N) iteration uses a new master stiffness matrix, so must be re-assembled and re-factorised before solving for each right hand side. The structure of the master stiffness matrix is always the same with the same finite element mesh and so the first three steps are performed only once. These six steps are now considered individually.

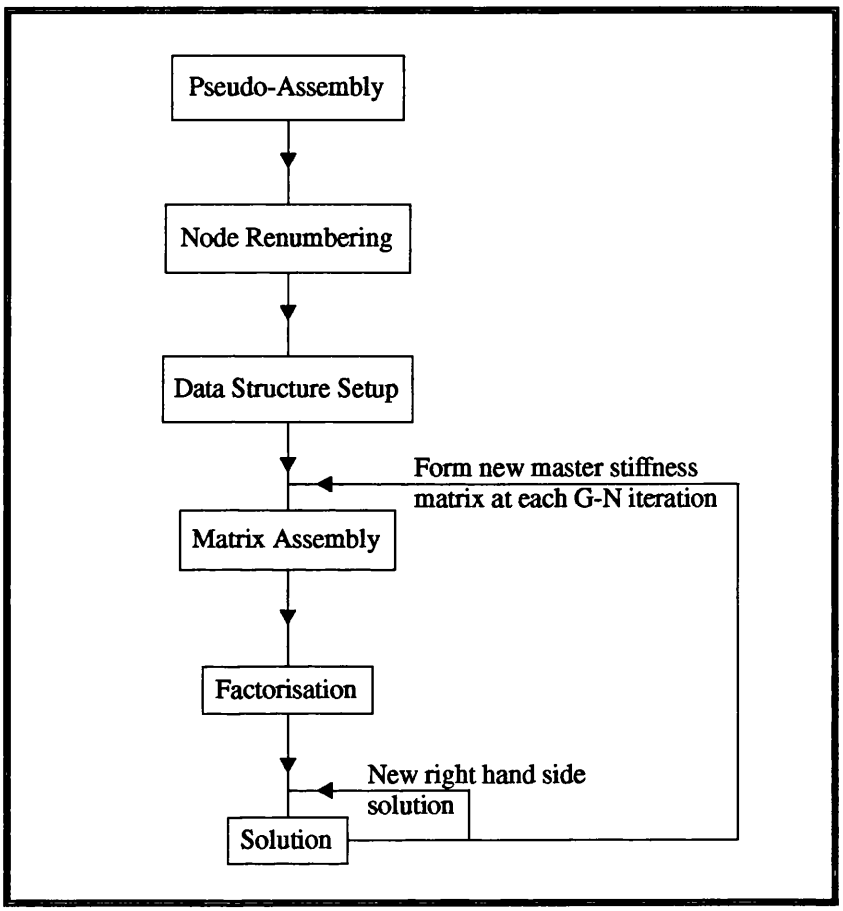


Figure 4.6 Steps to solution of sparse finite element equations

4.2.2.2.1. Pseudo-Assembly

The finite element assembly procedure produces the $r \times r$ master stiffness matrix **Y**. This matrix structure must be known for some of the node renumbering algorithms to be employed. The finite element assembly procedure is used solely to find the positions of the off-diagonal non-zero entries in the master stiffness matrix. This *pseudo-assembly* produces a one-dimensional array which has r segments, each corresponding to a column in **Y**. Each segment stores a list of the row numbers at which there is a non-zero

in the **Y** matrix. A pointer array pointing to the first non-zero element of each column in the one-dimensional array is also needed.

Various matrix entries are referenced more than once during the assembly procedure, so the pseudo-assembly implements a search ensuring that entries are not duplicated.

4.2.2.2.2. Node Renumbering

As mentioned, careful numbering of the nodes can reduce the amount of fill-in during Cholesky factorisation. Numerous node numbering schemes exist and a selection of these are considered in detail in Section 4.2.2.3.

4.2.2.2.3. Data Structure Set-up

Once the matrix re-ordering is chosen, the sparse data structures must be set-up so that the coefficient matrix, **Y** and the factorised matrix, **L** can be stored. The storage scheme is linked to the chosen factorisation and solution methods and are split as follows:

i) Envelope Methods
ii) General Sparse Methods

i) Envelope Methods

The band structure, which is examined in the Section 4.2.2.3, has predictable fill-in which facilitates the use of simple data structures to store such matrices. The variable band structure consists of rows of different lengths within the band, see Figure 4.5. The area within the band is also known as the envelope. The data storage structure used is the *envelope storage scheme*[54]. This uses a one-dimensional array to store all elements within the envelope. Matrix symmetry enables storing of the lower triangle elements only. The elements within the lower triangular envelope are stored row after row in the one-dimensional array.

The envelope storage scheme is also suitable for the factored matrix, **L** as factorisation fill-in only occurs within the envelope, see Figure 4.5. No fill-in position calculations are necessary. Setting up the data structures is merely a matter of finding the envelope size and setting up an array holding pointers to the start of each row in the one-dimensional array. The position of the first non-zero of each row in the factored matrix is readily obtained from the information produced in the pseudo-assembly stage.

ii) General Sparse Methods

The minimum degree and nested dissection orderings, which will be considered, produce a more scattered distribution of non-zero elements in the coefficient and factored matrices. A more complicated data structure is needed to access these elements. Linked lists[53] are often used, especially in general (non-symmetric) sparse matrices. However, the solvers used are obtained from George[54] and these need a data structure especially adapted to sparse symmetric factorisation and solution. This *compressed sparse scheme* uses a one-dimensional array to store the coefficient and factored matrices, along with the necessary pointer and element positioning vectors. Full details may be found in George[54].

The more general sparse orderings do not give the same factored matrix fill-in predictability that is associated with envelope structures (see Figures 4.8 to 4.10) . In order to find the positions of the fill-ins, a process known as symbolic factorisation must be performed. Symbolic factorisation simulates the numerical factorisation procedure in order to find the zero-nonzero structure of the factored matrix. The compressed sparse data storage scheme can be set-up simultaneously. George[54] provides an efficient routine to implement this process.

4.2.2.2.4. Matrix Assembly

In equation (4A.14), the master stiffness matrix, **K** is the coefficient matrix **Y**. Only the lower triangle of **Y** is stored in the appropriate one-dimensional data arrays, so just the lower triangle need be assembled as follows:

$$\mathbf{Y} = \mathbf{K} = \sum_{h=1}^{m} \sum_{i=1}^{4} \sum_{j=1}^{i} \left[\mathbf{k}_{\max(I,J)\min(I,J)} \right]_h \tag{4.8}$$

The *max(I,J)* and *min(I,J)* are functions to find the maximum and minimum of *I* and *J*. The last two summations translate into 2 computer program loops, both of which are very small. The loop overhead slows down the assembly process. The implementation time is decreased by replacing the 2 loops with 1 through the use of indirect addressing.

Once matrix **Y** in equation (4.3) has been assembled, the boundary conditions must be implemented before proceeding. The right hand side vector, **c** simply contains the nodal currents flowing into the mat, therefore, this vector holds the natural boundary conditions. These currents flow through the electrodes attached to the mat. Vector **c**

thus contains the electric current, in Amps, at row numbers which correspond to the electrodes and zero elsewhere. **Y** is singular, so a reference node is supplied by grounding a node, *k*. To reference this node, set the *k*th row and column of **Y** equal to zero and the main diagonal element to 1 and set the *k*th row of **c** to zero. This is the equivalent of implementing the rigid boundary conditions, $\mathbf{v}_k = 0$ into equation (4.3) i.e. grounding node *k*. The rigid boundary conditions must be implemented so that the assembled matrix **Y** becomes positive definite, before the finite element equations (4.3) can be solved.

4.2.2.2.5. Factorisation

The permutation matrix $\mathbf{PYP}^T$ is factored into the form $\mathbf{LL}^T$ using the Cholesky factorisation algorithm. The **Y** matrix is a positive definite symmetric matrix, which is a necessary condition for the Cholesky factorisation. Efficient implementation of the numerical factorisation algorithm is dependent on the storage scheme used. Appropriate library routines[54][55] optimise the numerical factorisations according to the data structure schemes used.

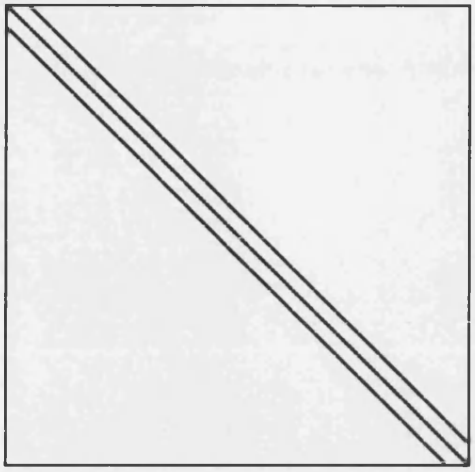
4.2.2.2.6. Solution

Equation (4.3) needs to be solved *p* times. The different solutions are for the different current patterns injected into the electrodes. The *p* different **v** vectors are each solved separately through forward and backward substitution. The factorisation of **Y** need only be performed once as the forward and backward substitution uses the same factorised matrix, **L**, for each of the *p* right hand side **c** vectors.

The vector **c** is also sparse in some situations and further computational savings may be made by taking advantage of these zeros. Formally, this is implementing sparse vector methods and is examined further in Section 4.2.2.4.

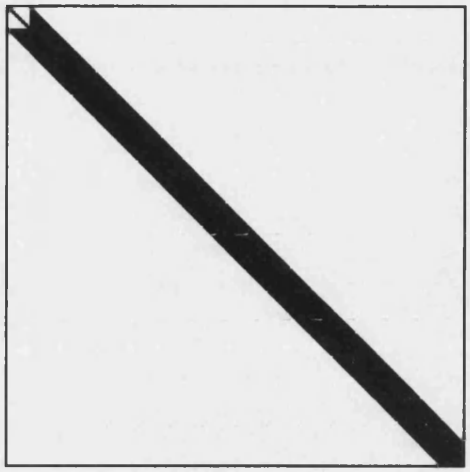
4.2.2.3. Sparse Matrix Re-orderings

The sparse matrix technique, in effect, finds the permutation matrix in equation (4.7). This corresponds to relabelling of the node numbers in the finite element mesh. Three different approaches to ordering are examined, namely; the envelope ordering, minimum degree ordering and nested dissection.



441×441 **Y** Matrix

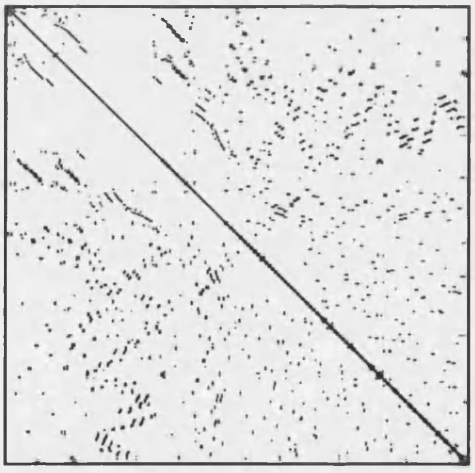
Non-zeros: 3721 Sparsity: 1.9%



441×441 $\mathbf{L}\backslash\mathbf{L}^{T}$ Matrix

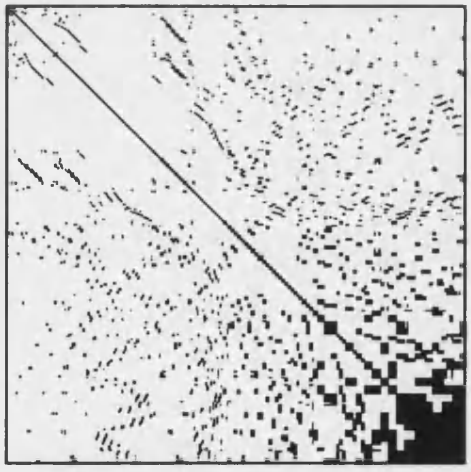
Non-zeros: 18921 Fill-in: 15200 (7.8%)

Figure 4.7 Matrix structure for the 400 element grid with the natural (nat) ordering



441×441 **Y** Matrix

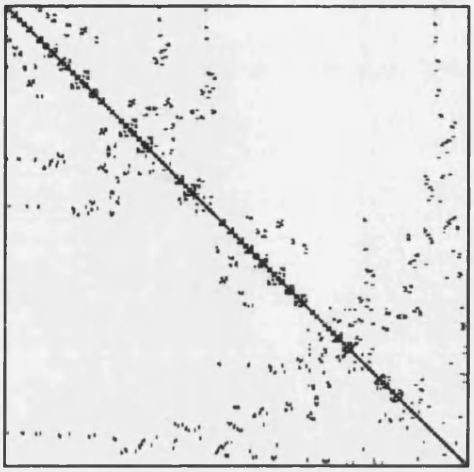
Non-zeros: 3721 Sparsity: 1.9%



441×441 $\mathbf{L}\backslash\mathbf{L}^{T}$ Matrix

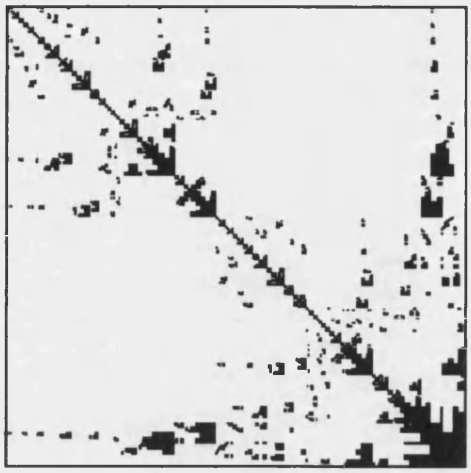
Non-zeros: 14075 Fill-in: 10354 (5.3%)

Figure 4.8 Matrix Structure for the 400 element grid with the minimum degree (MD) ordering



441×441 **Y** Matrix

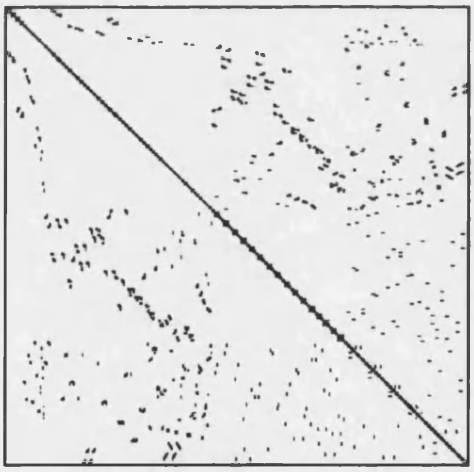
Non-zeros: 3721 Sparsity: 1.9%



441×441 $\mathbf{L}\backslash\mathbf{L}^T$ Matrix

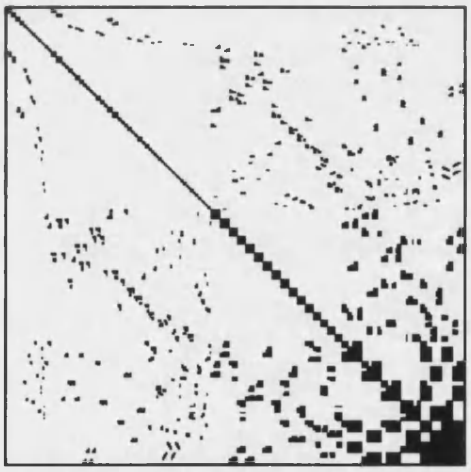
Non-zeros: 13461 Fill-in: 9740 (5.0%)

Figure 4.9 Matrix structure for the 400 element grid using the irregular nested dissection (IND) ordering



441×441 **Y** Matrix

Non-zeros: 3721 Sparsity: 1.9%



441×441 $\mathbf{L}\backslash\mathbf{L}^T$ Matrix

Non-zeros: 12735 Fill-in: 9014 (4.6%)

Figure 4.10 Matrix structure for the 400 element grid using the regular nested dissection (RND) ordering

4.2.2.3.1. Envelope Orderings

Matrices can be ordered so that all the non-zeros lie within an envelope around the main diagonal. During factorisation, fill-in occurs within this envelope and so envelope (also called skyline/profile/variable-band) algorithms operate entirely within this area and ignore all the zeros outside the envelope. Some algorithms ignore the variable bandwidths that make up the envelope and operate on a fixed bandwidth. The band structure and predictability of fill-in facilitates the use of the envelope storage scheme to store such matrices. Envelope solvers take advantage of this simple data storage scheme and as a result are quite fast on a per matrix element calculation basis.

The row by row or natural ordering of the mesh produces the narrow band structure of Figure 4.7 for the 400 element grid. The row by row ordering is produced by the mesh generator. Algorithms that can automatically order matrices into a band structure exist and the best of these is known as the Gibbs-Poole-Stockmeyer algorithm. The aim of this algorithm is to reduce the size of the envelope. This will in turn reduce the number of fill-ins and subsequent operation count during the factorisation and solution stages. An efficient implementation of the Gibbs-Poole-Stockmeyer algorithm by Lewis[56] was used in an attempt to improve the mesh generator's row by row ordering. It did not reduce the envelope size any further.

4.2.2.3.2. The Minimum Degree Ordering

The minimum degree algorithm is also known as the Tinney Scheme 2 ordering as it was first published as scheme 2 by Tinney[57] in 1967. Since then, the algorithm has been improved slightly and implemented more efficiently[58]. The interested reader is referred to this last reference for further details on the latest refinements to the minimum degree algorithm. The algorithm is specially tailored for a wide range of symmetric matrices and is a popular ordering for general sparse matrix solvers. In the next few paragraphs, the basic algorithm is presented along with an example of how the ordering is obtained when applied to a tiny finite element mesh. A basic knowledge of graph theory is assumed, but this can also be found in Appendix A1.

The minimum degree algorithm attempts to re-order the matrix for minimum fill-in during the factorisation stage of the Cholesky decomposition. It can be verified that when the graph has no loops (a tree graph), the minimum degree ordering introduces no fill-in. However, the algorithm does not necessarily always produce an ordering that gives minimum fill-in[53]. The graph for the finite element mesh used is shown in

Figure 4.11. This graph structure has loops and so cannot produce zero fill-in and the minimum degree algorithm cannot guarantee an ordering that will produce the minimum amount of fill-in. Tinney[57] describes the ordering as the numbering of the rows so that at each step of the process, the next row to be operated upon is the one with the fewest nonzero terms.

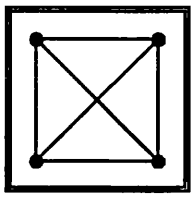


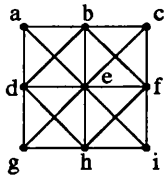
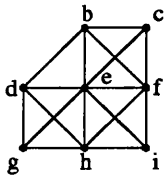
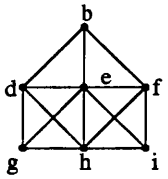
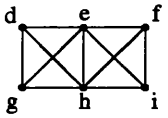
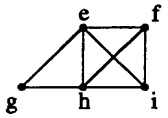
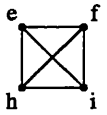
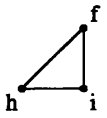
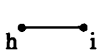
Figure 4.11 Graph of the single bilinear finite element

The minimum degree algorithm can be more explicitly stated through the use of graph theory:

1) Use the given symmetric graph, i.e. the unordered matrix structure as the current elimination graph.
2) Choose a node which has minimum degree in the current elimination graph. This node is numbered next. If more than one node has minimum degree, choose any node.
3) Remove this node from the current elimination graph to produce a new elimination graph.
4) This node elimination changes the degree of some uneliminated nodes so each node degree must be updated.
5) Repeat 2-4 above while there are still nodes that have not been eliminated (numbered).

An example of how the algorithm numbers a simple 4 element mesh is shown in Figure 4.12.

The Fortran algorithm used was obtained from George[54] and the exact implementational details are fully described in this book. The re-ordered 400 element mesh produces the coefficient and factorisation matrices shown in Figure 4.8.

Elimination Graph	Chosen Node	Node Number	Node Degree
	a	1	3
	c	2	3
	b	3	3
	d	4	3
	g	5	2
	e	6	3
	f	7	2
	h	8	1
•i	i	9	0

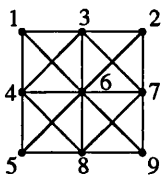
Resulting minimum degree ordering: 

Figure 4.12 Example of the minimum degree algorithm on a two element mesh

Tie-breaking

The minimum degree algorithm here uses no tie-breaking strategy, that is, when two or more nodes have equal minimum degree, no strategy is used to choose between them. The choice of node is in effect random as it depends entirely on the initial ordering. The example in Figure 4.12 always picks the first letter of the alphabet when there is a choice between two or more nodes of equal minimum degree. Different tie-breaking strategies induce different orderings and so the amount of fill-in and success of the minimum degree algorithm depends on the tie-breaking strategy used. On a square grid, Duff[53] presents an example where a row by row initial ordering produces less fill-in than a spiral (numbering nodes spiralling towards the middle) initial ordering. George[58] also found that a row by row initial ordering was best, so this is the one used. The same author suggests that the next significant advance of the minimum degree algorithm will probably be the development of effective and practical tie-breaking strategies in the selection of minimum degree nodes. This is currently an active area of research.

4.2.2.3.3. Nested Dissection Ordering

Nested dissection is an ordering algorithm which first had applications in finite element problems. The algorithm is adaptable to meshes comprising both regularly and irregularly shaped finite elements. The regular square finite element mesh is the relevant case, but an irregular algorithm may also be applied to a regular problem. The regular algorithm is examined in greater detail here.

George[59] and Duff[60] have published nested dissection algorithms especially tailored for regular square finite element grids. The latter publication is an improvement over the first in that it gives a superior ordering for the square $m \times m$ grid when m is not a power of 2. This superior algorithm is the one tested.

The nested dissection ordering algorithm hinges on dividing up the finite element mesh piece by piece. The mesh is separated into 4 subsets by a cross-shaped group of dividing nodes. The process is repeated on the remaining nodes in each subset until each subset can no longer be subdivided. The dividing group of nodes, or dissection sets, in each subset are numbered last. Thus, the mesh is repeatedly divided by the dissecting sets and hence the term 'nested dissection' is coined. The resulting 4 dissection sets for the 100 element grid are shown in Figure 4.13, with a few sets shaded for clarification. The nested dissection algorithm numbers the nodes in set 1 first, followed by those in sets 2,3... and so on.

Methods for numbering the nodes within each dissecting set can be optimised and was studied by Duff[60]. He found the minimum degree ordering to be best. This node numbering applied to the 100 element mesh using the dissecting sets of Figure 4.13 is shown in Figure 4.14.

The dissection sets were obtained by coding the algorithm of Duff[60]. The minimum degree numbering of the nodes within each dissection set was done by an adaption of the Fortran algorithm in the book by George[54].

The non-zero matrix structure for the 400 element mesh is shown in Figure 4.10. As can be seen, the non-zeros are irregularly scattered throughout the matrix. The compressed sparse scheme is best for storing such a scattered matrix structure.

George[61] developed an automatic general purpose nested dissection ordering which can also be applied to irregular finite element problems. The authors claimed that the algorithm gives an efficient ordering for regular square finite element grids. The Fortran code for this algorithm was readily available in the book[54] by the same authors and the resulting ordering for the 100 element mesh case is shown in Figure 4.15. Comparisons with the regular mesh ordering of Figure 4.14 show that the dissecting sets are not quite the same. The effects of not using the theoretical regular orderings for a square mesh are examined in the Section 4.2.2.5.

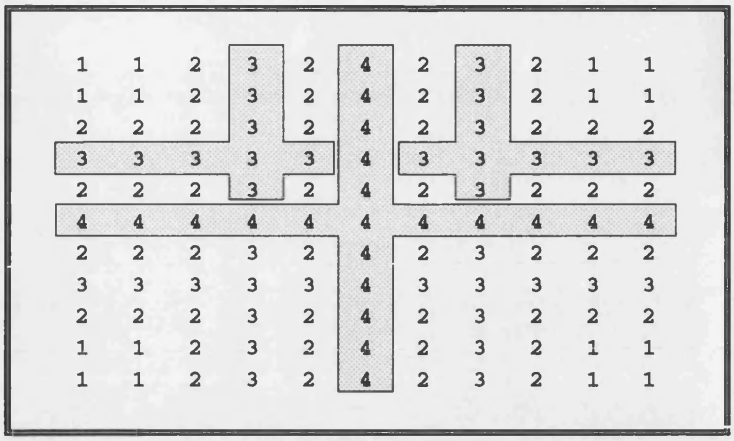


Figure 4.13 Regular nested dissection sets for the 100 element mesh. The 4th dissection set and 2 of the 3rd dissection sets are shaded.

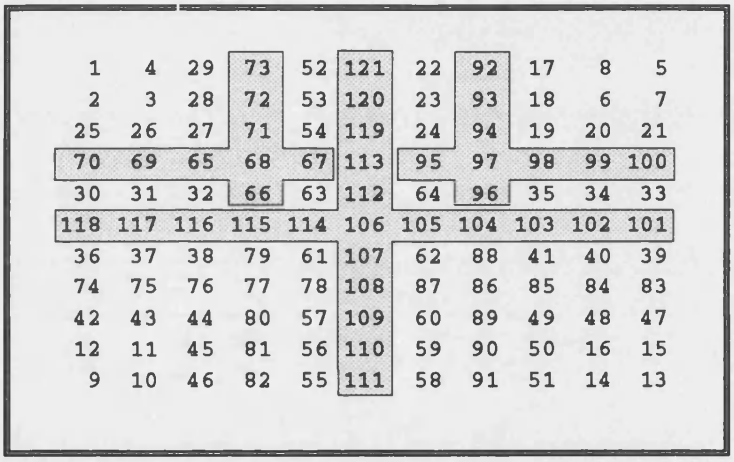


Figure 4.14 Regular nested dissection ordering for the 100 element mesh using the dissection sets of Figure 4.13. The shaded sets are also the same.

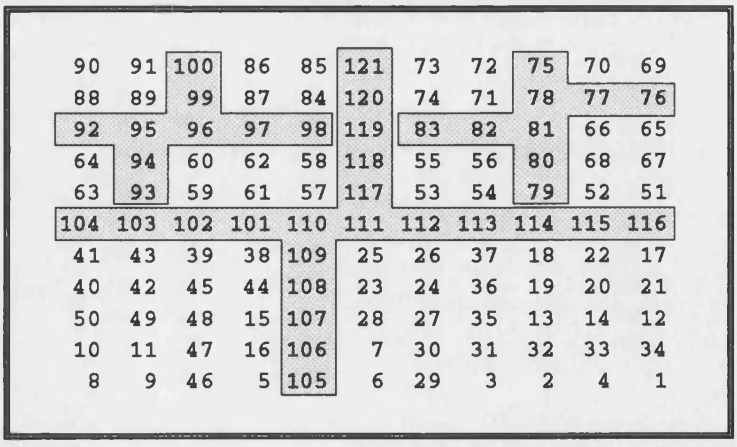


Figure 4.15 Irregular nested dissection (automatic algorithm) ordering for the 100 element mesh. A few groups within the dissection sets are shaded for distinction.

4.2.2.4. Sparse Vector Methods

The incentive for using sparse vector methods[62] is to improve on the sparse matrix method by further reducing the number of unnecessary matrix operations. Achievement of this is possible in two instances; when the right hand side vector is sparse and/or only a subset of the solution vector needs to be known.

In the first instance, the forward substitution can be optimised by using a subset of the columns of the factorised matrix, **L**, and is known as fast forward (FF). Similarly in the second instance, a subset of the rows of $\mathbf{L}^T$ speed up backward substitution and is known as fast back (FB). The ordered list of columns in **L** for FF, or rows in $\mathbf{L}^T$ for FB, which are absolutely necessary for solution, define what is known as the sparse vector path. The path length in FF and FB is the number of these necessary columns in the ordered list. The path length of full forward and backward substitution is equal to the total number of columns in **L**.

The paths for FF and FB are usually different and are respectively, functions of the sparsity structures in the right hand side and solution vectors. The paths are also functions of the sparsity structure in **L**, which in turn is a function of the **Y** matrix ordering strategy used. The most important task for sparse vector methods is to find this path in an efficient manner. Tinney[62] explains methods to do this and how to implement FF and FB in a clear concise way. NAG[55] and George[54] provide library routines which are used for full forward and backward substitution. These were then modified, when appropriate, for implementation of FF and FB.

4.2.2.4.1. Voltage Solution

In the application of EIT to measuring pressure from a mat, a current source is used for current injection into and extraction from the mat. Each right hand side column vector **c** in the finite element equations only has non-zeros in positions which correspond to the electrode numbers used by the current source. In single current injection, current is injected into one electrode and extracted from one other and so **c** has two non-zeros. This is a prime candidate for using FF as **c** is highly sparse. When optimal currents are used, each and every electrode has an injected current. In this case, **c** is not particularly sparse. Additionally, the electrodes are not topologically clustered as they are spread fairly well all over the mat. Sparse vector methods (FF) are not considered in this case as these conditions are not ideal for using them[62]. The Gauss-Newton method needs all

voltages on all nodes for the Jacobian matrix calculation so full backward substitution is necessary.

4.2.2.4.2. Matrix Inverse Solution

Another situation arises where very similar equations must be solved. This is in the implementation of the Gauss-Newton method where some columns of the **Y** inverse matrix are needed, see Section 5.4.1 for details. Large computational savings can be made by using the same factorised matrix, **L**. Forward and backward substitution are then used to solve for the inverse, $\mathbf{Y}^{-1}$, by using the appropriate column of the identity matrix, **I** as the equation right hand side:

$$\mathbf{Y}\mathbf{Y}^{-1} = \mathbf{I} \tag{4.9}$$

Each column vector in **I** is sparse with just one non-zero. Such a sparse vector is called a singleton by Tinney[62] and is the best condition for solution optimisation by FF. The Gauss-Newton method needs all the elements in only e columns of $\mathbf{Y}^{-1}$, where e is the number of nodes representing electrodes. Symmetry in **Y** enables the solution of all the elements in the e rows as an alternative. For a 400 element finite element mesh, e is approximately 7% of the total number of columns in **Y**.

The two ways of calculating the required elements in $\mathbf{Y}^{-1}$ are thus by rows or by columns. Calculating each column involves FF followed by full back substitution for the e electrode columns only. Calculating e rows requires FF and FB for every column in $\mathbf{Y}^{-1}$. FB would not give particularly significant improvements over full back substitution for the same reasons mentioned in Section 4.2.2.4.1. The obvious choice is then calculation by columns as only e FFs need be performed.

4.2.2.5. Choice of Solution Method

This section studies the node ordering and sparse vector method combinations for minimising the time to equation solution. The chosen method must be optimal for both the voltage solution and inverse matrix solution.

4.2.2.5.1. Timing Results

Timing minimisation is the final goal, but the CPU timing results taken vary by approximately 10% due to the multitasking environment used. To aid comparisons, the number of operations, i.e. the number of multiplications and divisions, is used in addition

to timing information. The majority of the time spent in solving the EIT forward problem related equations is in the factorisation and solution stages, as Figure 4.16 shows. The time to complete these two stages need further examination as the remaining stages use negligible CPU time. All timing results in this section were taken on a Sun 10/41 and averages taken to compensate for fluctuations.

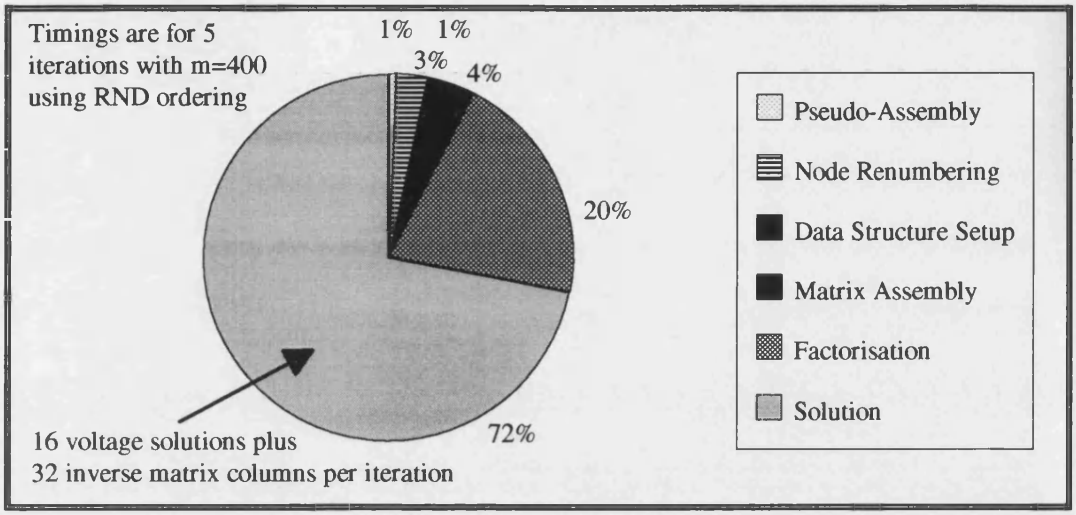


Figure 4.16 Typical timing breakdown for voltage and inverse calculation in the Gauss-Newton method

The factorisation times are dependent on the data storage scheme and the number of operations performed. Figure 4.17 compares the number of operations executed during Cholesky factorisation. In this and subsequent tables, **Nat** is the natural or row by row ordering, **MD** is the minimum degree ordering, **IND** is the ordering from the irregular/automatic nested dissection algorithm and **RND** is the regular square mesh nested dissection ordering. The mesh size is represented by m, where m is the number of elements in the square mesh. The most relevant results from this table are for the larger grid sizes as they provide a more accurate representation of the physical model. The nested dissection algorithms have superior operation counts and the algorithm specially tailored for the regular mesh (RND) is better than the more general irregular algorithm (IND). The IND algorithm still outperforms the MD algorithm for the larger meshes. The operation count improvements are bigger for the larger meshes. The amount of fill-in has a direct effect on the operation count. This is apparent in the percentage of fill-in shown with the matrix structures of Figures 4.7 to 4.10.

The solution stage produces similar results. The full backward substitution and full forward substitution / singleton FF count can be seen in Figure 4.18. The full forward substitution is applicable to the voltage calculation, but FF is used in the $\mathbf{Y}^{-1}$ calculation. The benefits of using FF are attributable to the shorter path lengths. The average singleton path length for calculating all columns in the $\mathbf{Y}^{-1}$ matrix is shown in Figure 4.19. The row labelled **Full** is not a singleton path length, but the full path length, that is, the path for full forward substitution. The natural ordering is not particularly suitable for FF as once the first right hand side non-zero is encountered, the path is full from then on. The more scattered non-zero distributions in the minimum degree and nested dissection ordered matrices lead to path lengths much shorter than that of the full path. These are more suitable for sparse vector methods. The effect of these FF path lengths is reflected in the operation counts of Figure 4.18.

m	25	64	100	256	400
Nat	990	4 496	9 580	50 336	113 560
MD	802	3 550	7 574	43 781	85 041
IND	936	4 232	8 546	35 929	71 667
RND	808	3 460	7 005	32 282	64 987

Figure 4.17 Operation count for Cholesky factorisation

m	25		64		100		256		400	
	Full	FF	Full	FF	Full	FF	Full	FF	Full	FF
Nat	492	365	1 602	1 195	2 882	2 153	10 370	7 763	19 362	14 503
MD	436	288	1 366	888	2 428	1 502	8 590	5 391	14 516	8 661
IND	472	296	1 506	903	2 606	1 543	8 084	4 671	13 902	7 947
RND	438	281	1 358	832	2 348	1 425	7 618	4 471	13 176	7 649

Figure 4.18 Average operation counts comparing full forward substitution and singleton FF. Operation counts include full backward substitution.

m	25	64	100	256	400
Full	36	81	121	289	441
Nat	18	41	61	145	221
MD	12	22	26	52	59
IND	10	17	23	39	50
RND	11	18	24	41	54

Figure 4.19 FF Average singleton path lengths

The RND ordering produces the lowest operation count for both FF and full forward substitution. This optimal ordering and the implementation of sparse vector methods is also reflected in the timing comparisons for one Gauss-Newton iteration, i.e. calculating e columns in $\mathbf{Y}^{-1}$ and p voltage columns. Figure 4.20 analyses these timing comparisons

for the two larger grid sizes. The times shown use the full forward substitution for the voltage calculation, as would be the case with optimal currents. The smaller grid size timings are not considered as their modelling inaccuracies are not ideal for use in the Gauss-Newton method. The smaller grid size trends can be obtained from the operation count tables. Also included in this table for reference, is the time taken when not taking advantage of any sparse methods, i.e. using a dense matrix solver from NAG[55]. When using the largest grid, the dense solver is approximately 24 times slower than the general solver with RND ordering and indicates the relative advantage of using the best sparse method over standard dense methods.

The time to solution is dependent on the data storage schemes used by the different sparse solvers. This is demonstrated by comparing the envelope and general sparse solvers with an ordering (Nat) that they can both operate on. The benefits of using the simpler storage scheme are especially evident in the factorisation times but the solution times are very similar. However, when the envelope solver is compared with the general solver using the RND ordering, the envelope solver is slower. Another envelope solver from NAG[55] was implemented, but proved to be much slower than the aforementioned one.

The timing results already presented calculate all elements in e columns in the $\mathbf{Y}^{-1}$ matrix. The alternative method of calculating e elements in all columns, using FB, was also studied. The FB path length was approximately half that of the full path and considering this had to be done for all r columns, where the ratio $r : e$ is ~14 for the large mesh, the solution times are much greater. Another strategy is to number the electrode nodes last and to use the minimum degree ordering on the non-electrode nodes. This approach degrades the minimum degree algorithm's low fill-in, but gives the advantage that only the first e backsubstitutions need be performed, i.e. a FB path length of only e. This ordering was implemented, but once again did not give very impressive results despite the short FB path length (0.45s for one G-N iteration, $m = 400$). This is because of the extra matrix fill-in, but largely due to FB being performed on all r columns.

	m	256	400
Dense Solver (Any ordering)	Factorisation	0.21900	0.69600
	Full Solution	0.01480	0.05370
	Total	0.929	3.274
Envelope Solver (Nat)	Factorisation	0.01230	0.03510
	Full Solution	0.00233	0.00475
	FF Solution	0.00159	0.00328
	Total	0.100	0.216
General Sparse Solver (Nat)	Factorisation	0.01820	0.04490
	Full Solution	0.00266	0.00472
	FF Solution	0.00166	0.00334
	Total	0.114	0.227
General Sparse Solver (MD)	Factorisation	0.01570	0.04060
	Full Solution	0.00201	0.00380
	FF Solution	0.00118	0.00203
	Total	0.086	0.166
General Sparse Solver (IND)	Factorisation	0.01360	0.03570
	Full Solution	0.00196	0.00359
	FF Solution	0.00100	0.00184
	Total	0.077	0.152
General Sparse Solver (RND)	Factorisation	0.01200	0.03050
	Full Solution	0.00186	0.00340
	FF Solution	0.00092	0.00166
	Total	0.071	0.138

Notes:
Total time is that for 1 Gauss-Newton iteration, i.e.
Total = Factorisation + Full Solution * p + FF Solution * e,
where e = 32 & p = 16. Times are in seconds on a Sun 10/41.
The dense solver does not implement FF.

Figure 4.20 Timing comparisons for different solvers and different orderings

4.2.2.5.2. Discussion

Many different ordering strategies have been compared with each other in the various sparse matrix literature. The fastest solutions depend to a large degree on the matrix sizes, the original ordering and the degree number of each node i.e. the connectivity between nodes. Both the way the algorithms are coded and the type of computer used

also have an effect. Consequently, it is difficult to foresee which matrix ordering method would be the best without testing the different ordering strategies.

The regular nested dissection algorithm proved to be the best ordering for minimising the equation solution times. Better tie-breaking strategies could be found to improve the present minimum degree algorithm. Tie-breaking strategies are currently being researched, but the nested dissection ordering is chosen as it can be viewed as a minimum degree ordering with a perfect tie-breaking strategy when applied to regular grids[58]. Another reason for the success of the nested dissection algorithm is that it is a global ordering strategy as the most important decisions are made on the entire matrix, whereas the minimum degree algorithm is a local ordering strategy[53].

A paper by Gomez[63] suggests three new matrix re-ordering algorithms to enhance sparse vector performance. The new algorithms are compared to the minimum degree ordering and are in effect tie-breaking strategies for it. The best algorithm shows an up to 20% improvement in the number of operation counts and path lengths for the test matrices used. The nested dissection algorithm with FF exhibits comparable improvements over the standard minimum degree ordering. For example, from Figures 4.19 & 4.18, the path length for the 256 element model is 21% shorter and 17% fewer arithmetic operations are needed. The algorithms in this paper also produce greater fill-in than the minimum degree, which results in longer factorisation times, whereas the nested dissection has a lower fill-in than the minimum degree algorithm.

The data structures used by the minimum degree algorithm and nested dissection algorithm are complicated due to the spread of non-zeros throughout the matrix. The envelope solvers access matrix elements faster due to the simpler data structures used. If the number of operations to be performed by an envelope ordering is similar to that of the general sparse solvers, then the envelope methods are quicker. However, in the situation studied here, the number of arithmetic operations required by the envelope ordering is greater and the net result is a faster solution time using the more complicated matrix orderings. The envelope orderings cannot take great advantage of sparse vector algorithms either.

George[64] and Duff[60] also showed that the nested dissection ordering is best for finite element meshes similar to that used for the pressure mat, but neither considered sparsity in the right hand side vector. One of the orderings not yet mentioned that George[64]

made comparisons against is the so called one-way dissection. The one-way dissection ordering is a compromise between band and nested dissection orderings. It dissects the mesh so that the dissected blocks may be stored and solved by envelope methods. However, it was shown to give poor factorisation times so was not considered.

Woo[65] considered envelope ordering methods and the three Tinney[57] schemes in his EIT system. He suggested using the minimum degree ordering, but this was for a circular finite element mesh comprised of triangular elements.

The solution methods considered are direct methods especially adapted to symmetric sparse matrices. Alternatively, iterative methods could be used, but were not considered as they are best for single solutions, are best for larger problems and are more efficient when a good starting vector is available.

Both Woo[65] and Paulson[50] have reduced EIT reconstruction times through implementation on parallel computers. Woo argues that the algorithms are easily parallelisable. However, Paulson found the porting of the algorithms to a parallel environment a formidable task; the implementational effort required was not worth the gain in speed considering the rapid rate of scalar processing power improvement. He recommends waiting for better tools for parallel implementation as parallel code is presently not portable and is difficult to maintain or modify. The operation counts shown in this chapter are a direct measure of processor time required when using the general sparse solver on a scalar computer, the Sun 10/41. The choice can be made largely on the operation count, as was shown in the timing results, but this would not necessarily be the case in a parallel environment.

4.3. Conclusions

The time that one Gauss-Newton iteration takes has been reduced on a superscalar computer. In real-time applications, a reduction in reconstruction time enables better resolution in the final reconstruction pressure distribution as more elements may be used in the FEM mesh within a given time limit. The areas targeted for optimisation were in the forward solution and the calculation of the Jacobian matrix. These equation solutions are suitable for sparse matrix optimisation. The nested dissection algorithm is the best node renumbering strategy for a regular rectangular mesh comprised of square bilinear elements. Additionally, the nested dissection ordering was found to produce a singleton path length lower than that produced by the minimum degree ordering. These low path

lengths further reduce the time to calculate the Jacobian matrix. The suitability of a mesh comprised of bilinear elements for optimisation by sparse methods, is one factor in choosing them to model the mat. Before the suggestions in this chapter were implemented, the total time spent in the forward problem and Jacobian matrix calculation dominated the reconstruction of the pressure distribution. Now it is a minor part of the Gauss-Newton inverse problem solution time.

5. Optimisation Techniques for Solving the Inverse Problem

The inverse problem is classified as an optimisation problem, for which there are many different algorithms for solution. This chapter examines the suitability of many such algorithms and narrows the choice down to a select few for closer study. The final part of this chapter analyses efficient methods for the formation of the Jacobian matrix required by the optimisation algorithms.

Methods for solving a problem by optimisation fall naturally into two different classes[66]. The first class includes those based on a direct search and use function evaluations only. The second class consists of a collection of gradient methods which require gradient information in addition to function evaluations. The gradient methods include a special class when the function being minimised has a certain structure; namely, a sum of squares. The least squares approach, which uses a sum of squares, is particularly relevant to solving the EIT inverse problem and is studied in depth. Firstly, the suitability of direct search methods for solving the inverse EIT problem is examined.

5.1. Reconstruction by Direct Search Methods

Direct search methods use a strategy to minimise the objective function, which is based on the comparison of values of the objective function without any use of the derivatives. Most direct search methods have been developed from heuristic approaches, such as directing the search from results of conducted explorations down a set of defined independent direction vectors. Other approaches are based on the use of statistical designs. The primary interest in direct search methods is their generality and since they usually make no more assumptions other than continuity of the function, they can be applied to a very wide class of problems. They are particularly useful for cases where the function is non-differentiable and where the first partial derivatives of the function are discontinuous[67]. These conditions can cause difficulties with the more theoretically based gradient methods and so direct search methods can be very useful in these instances.

Since the problem considered in this thesis does not include any of the aforementioned criteria and since direct search methods often have a slow rate of convergence[67], the more theoretically based gradient methods are preferable.

5.2. Reconstruction by the Least Squares Method

This section considers various least squares approaches to reconstruct the pressure distribution. The theory behind each approach is presented and explained in this chapter. The results using a computer simulation of the system are presented in the following chapter.

5.2.1. General Introduction

Knight[1] showed that the Gauss-Newton approach for solving the resistivity distribution in EIT is very successful. This approach is just one least squares optimisation technique to solve the forward problem. In this chapter a number of different least squares approaches are considered to improve the reconstruction. The least squares approach uses a sum of squares and is a special case of gradient optimisation methods. When the problem can be formulated in such a special way, it may or may not be better to use specific methods to solve it. The arguments for choosing a least squares method over a more general gradient method are discussed.

Numerical methods that attempt to find a minimum or maximum of any given nonlinear objective function are generally iterative and start from an initial estimate for the solution. Each iteration generates an improved estimate over the previous one. The iterative procedure is halted upon fulfilling some criteria indicating that no further improvements can be made. Different procedures are characterised by their approaches to each iterative improvement and the least squares approach is one such iterative procedure.

In the least squares approach, the objective function, Φ is defined as a sum of squares of nonlinear functions. In our case the objective function is the sum of the equally weighted mean square difference between the measured and estimated voltages[46] and is represented by the equation

$$\Phi = \tfrac{1}{2}\sum_{i=1}^{n}\left[\mathbf{f}(\rho)-\mathbf{v}_o\right]_i = \tfrac{1}{2}\left\|\mathbf{f}(\rho)-\mathbf{v}_o\right\|_2^2 = \tfrac{1}{2}\left[\mathbf{f}(\rho)-\mathbf{v}_o\right]^T\left[\mathbf{f}(\rho)-\mathbf{v}_o\right] \qquad (5.1)$$

where $\mathbf{f}(\rho)$ is defined as a function mapping an m-dimensioned resistivity distribution vector into an n-dimensioned voltage observation vector. The voltage observation vector is defined as $\mathbf{v}_o$ and contains the measured voltage readings at electrodes for sets

of applied current patterns. The resistivity distribution vector corresponding to the discrete impedance distribution is defined as ρ. For any estimated ρ and applied current pattern, a corresponding set of voltages, that is, vector $\mathbf{f}(\rho)$, can be deduced from the finite element model of resistivity distribution.

The ½ in the minimising function of equation (5.1) is included merely to avoid the appearance of a factor of two in the derivatives. The residuals are the elements $[\mathbf{f}(\rho)-\mathbf{v}_o]_i$ and the least squares approach attempts to minimise the vector of residuals and does so by suitable selection of a resistivity distribution. So, when the two voltage distributions are close, the estimated resistivity distribution is accepted to be the desired one. The final resistivity distribution that minimises Φ in equation (5.1) is denoted ρ^*. This is then the resistivity distribution that corresponds to the pressure distribution being sought.

The function $\mathbf{f}(\rho)$, is actually not just a function of the resistivity distribution, but also of the current distribution and electrode location. The last two are chosen and are thus known. Optimal choice of these two is considered later in this thesis.

Equation (5.1) is an unweighted least squares minimising function. A weighted least squares EIT function has been considered by Woo[68]. He found that the weightings amplified contrasts in conductivities, but gave a less accurate reconstruction.

The least squares problem defined in equation (5.1) can be minimised by a general optimisation method, but the properties of equation (5.1) make it worthwhile to use methods designed specifically for the least squares problem[69]. These properties are shown in the following section on the Gauss-Newton method.

5.2.2. The Gauss-Newton Method

The well known Gauss-Newton method was first used in Electrical Impedance Imaging applications by Yorkey[40][46], who also termed it the modified Newton-Raphson Method. The Gauss-Newton method is the generalised least squares method, upon which the methods following this section are based. A brief derivation of the Gauss-Newton method follows where much of the terminology used is introduced.

To find the only unknown in the forward problem, ρ^*, the objective function is minimised using standard mathematical procedures. That is, differentiating the objective

function with respect to $\boldsymbol{\rho}$ and setting the result to the zero vector, **0**. This gives the gradient vector, **g**

$$\mathbf{g} = \Phi'(\rho) = \mathbf{J}^{\mathrm{T}}[\mathbf{f} - \mathbf{v}_0] = \mathbf{0} \tag{5.2}$$

The term **J** is known as the Jacobian matrix and is an $n \times m$ matrix defined by

$$\mathbf{J}_{ij} = \frac{\partial \mathbf{f}_i}{\partial \rho_j} \tag{5.3}$$

To linearise equation (5.2), as it is still a nonlinear function of $\boldsymbol{\rho}$, a Taylor series expansion of equation (5.2) about an arbitrary point, $\boldsymbol{\rho} = \boldsymbol{\rho}^k$ at iteration k is taken i.e.

$$\Phi'(\rho^k + \Delta\rho^k) \approx \Phi'(\rho^k) + \Phi''(\rho^k)\Delta\rho^k \tag{5.4}$$

where

$$\Delta\rho^k = \rho - \rho^k \tag{5.5}$$

The term Φ'' is known as the Hessian matrix. Using the chain rule, the Hessian matrix is given by

$$\Phi''(\rho) = \mathbf{J}^{\mathrm{T}}\mathbf{J} + \mathbf{B} \tag{5.6}$$

Yorkey[40] gives **B** explicitly as

$$\mathbf{B} = [\mathbf{f}'']^{\mathrm{T}}[\mathbf{I}_m \otimes [\mathbf{f} - \mathbf{v}_0]] \tag{5.7}$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\otimes$ is the Kronecker matrix product[70]. These last two equations show that the Hessian of the least squares problem consist of a special combination of first and second order matrices. Yorkey[40] stated that **B** is difficult to calculate explicitly and is negligible relative to $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ and therefore the Hessian matrix is approximated by

$$\Phi''(\rho) \approx \mathbf{J}^{\mathrm{T}}\mathbf{J} \tag{5.8}$$

This approximation is only valid within some small region about $\boldsymbol{\rho}^*$ and Yorkey[46] found this to be valid with his models. In fact, least squares methods are typically based on the premise that eventually, during the iterative process, the first order term $\mathbf{J}^T\mathbf{J}$ will dominate the second order term[69] in equation (5.6) and so the second order terms are neglected. This dominance occurs when the residuals at the solution are small.

Substituting equations (5.2) and (5.8) into (5.4), setting (5.4) to $\mathbf{0}$, the Gauss-Newton equation is obtained:

$$\left[\left[\mathbf{J}^k\right]^T\mathbf{J}^k\right]\Delta\boldsymbol{\rho}^k = -\left[\mathbf{J}^k\right]^T\left[\mathbf{f}^k - \mathbf{v}_o\right] \tag{5.9}$$

To solve for $\Delta\boldsymbol{\rho}^k$, the approximated Hessian must be non-singular. Equation (5.9) defines the iterative procedure to estimate the final resistivity distribution $\boldsymbol{\rho}^*$. At iteration k, $\boldsymbol{\rho}^k$ is updated from equation (5.9) to give

$$\boldsymbol{\rho}^{k+1} = \boldsymbol{\rho}^k + \Delta\boldsymbol{\rho}^k \tag{5.10}$$

A stopping criterion which indicates that $\boldsymbol{\rho}^{k+1}$ is the best estimate for the resistivity distribution halts the iterative procedure, whereupon, it is said to have converged. Convergence criteria are considered later in this section.

The vector $\Delta\boldsymbol{\rho}^k$ that solves equation (5.9) is called the Gauss-Newton direction and the method in which this vector is used as a search direction is the Gauss-Newton method. A diagrammatic summary of the Gauss-Newton method is shown in Figure 5.1. The Gauss-Newton procedure converges and can achieve a quadratic rate of convergence when the size of the second order term in equation (5.6) tends to zero[69].

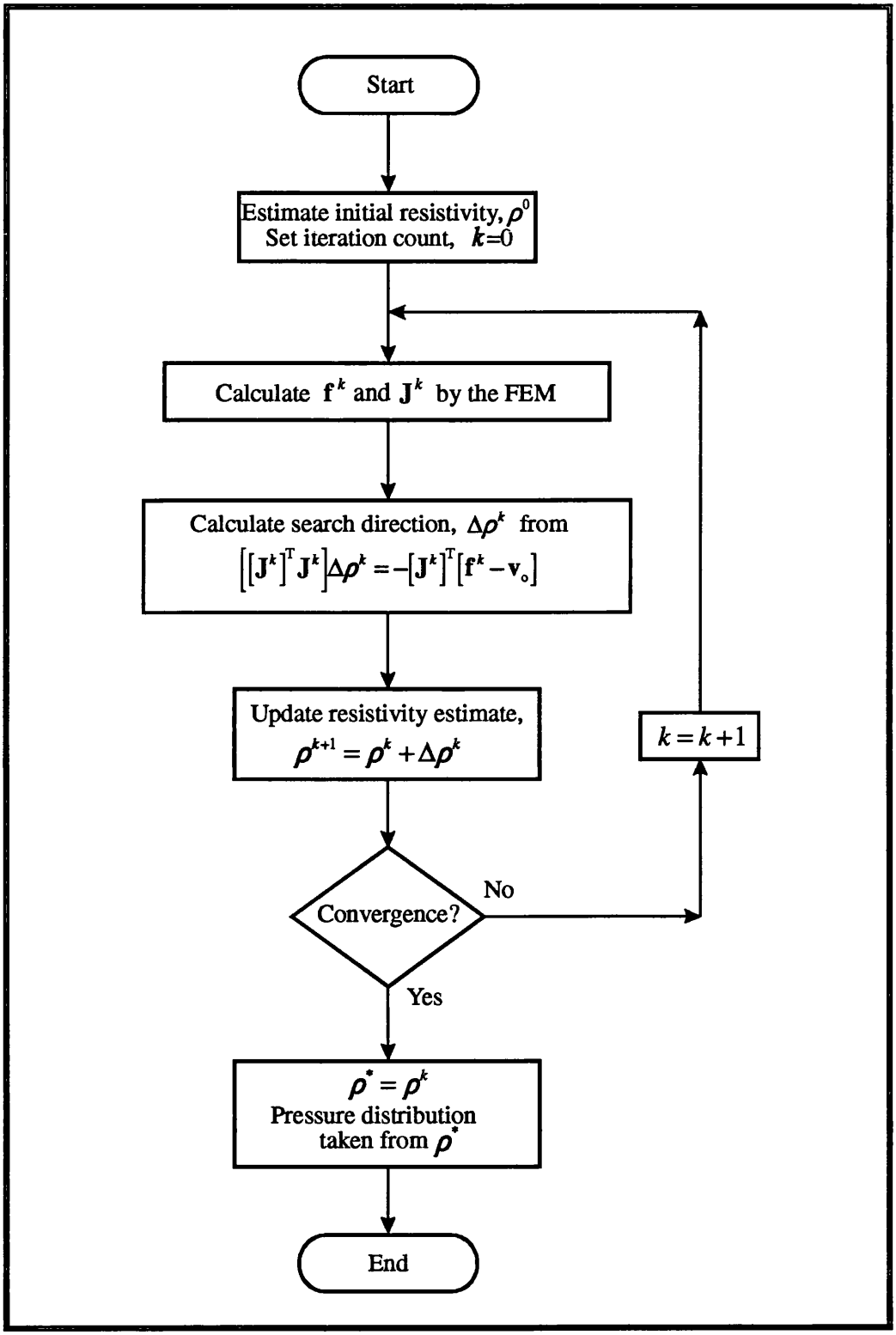


Figure 5.1 Flow Chart for the Gauss-Newton Method

5.2.3. The Damped Gauss-Newton Method

The size of the step to proceed in from the current point, in an attempt to find a better approximation to the resistivity vector, need not be unity as defined by equation (5.10). The size, α^k, of this direction is termed the step length and the updating equation (5.10) can be changed to incorporate the step length,

$$\rho^{k+1} = \rho^k + \alpha^k \Delta\rho^k \tag{5.11}$$

where $0 < \alpha^k \leq 1$. This then defines the damped Gauss-Newton Method[71]. The step length is chosen so that the function being minimised is reduced along the search direction, $\Delta\rho^k$, at each iteration, i.e.

$$\Phi^{k+1} < \Phi^k \tag{5.12}$$

Satisfying this equation by selection of the step length can then prevent divergence that might occur with a unit step length. Computation of the step length could be optimal in that it finds a minimum along the search direction, or it could be acceptable in that equation (5.12) is satisfied by many different step lengths. Searching for a minimum along the search direction requires greater computation per iteration, but usually fewer iterations will be needed for convergence. However, the total computation is usually less if an acceptable value, as opposed to an optimum value, for the step length is used[71].

5.2.4. The Gauss-Newton Method with Second Derivative Approximations

The assumption that the Hessian's first order term will dominate the second order term, is not always justified. The Gauss-Newton method ignores this second order term, **B** in equation (5.7). As mentioned, this matrix is difficult to calculate explicitly, but there are methods which approximate matrix **B**.

Before looking at a method to approximate **B**, a definition is needed of when these second order terms are necessary. Loosely, this is when the residuals of the Hessian at the solution are "large". It can be seen from equation (5.7) that when the residuals, $[\mathbf{f} - \mathbf{v}_o]$ are not zero at the solution, **B** can no longer be neglected. Gill[69] gives one of a few definitions in the available literature for a large residual problem as one in which the optimal residual exceeds the small eigenvalues of $\mathbf{J}^T\mathbf{J}$. When the optimal residual is so

large that it is comparable to the largest eigenvalues, Gill also suggests that no advantage is to be gained by exploiting the least-squares nature of the objective function and one may as well use a more general minimisation method. It must be noted that the residuals are large when the errors in the electrode voltage measurements are large. It was found that the appropriate eigenvalues indicate that the problem is a large residual one, but not too large to abandon the use of the least-squares objective function. This was for expected practical voltage errors in the region of 1%. Methods to approximate **B** need to be studied.

A method to approximate the second order derivatives is based on using a quasi-Newton technique. This is the suggestion of Gill[72]. Quasi-Newton methods, also called variable-metric or secant methods by some authors, are a large class of general optimisation solution methods which are frequently used to solve general unconstrained minimisation problems. They are usually used to approximate the entire Hessian matrix and not just the second derivatives of the Hessian matrix. They were not considered for the entire Hessian matrix approximation for the reasons given in the previous paragraph. These same arguments hold for conjugate-gradient methods, which are closely related to quasi-Newton methods[69] and are another popular method for solving optimisation problems. Conjugate-gradient methods generate search directions without storing any matrices and so are particularly suited to very much larger problems. Artola[73] recently evaluated the BFGS quasi-Newton method on an EIT problem. Results comparable to the Gauss-Newton method were shown just for zero measurement error, a small residual case. Unfortunately, these are unlikely to be replicated in the large residual conditions of voltage measurement error.

The quasi-Newton method then attempts to construct the second derivative matrix by gathering information obtained from previous iterations. Given $\mathbf{H}^k$, the kth quasi-Newton approximation to $\mathbf{B}^k$, the updated approximation for the next iteration, $\mathbf{H}^{k+1}$ is obtained by adding a correction or update matrix, $\mathbf{U}^k$ to $\mathbf{H}^k$ as follows

$$\mathbf{H}^{k+1} = \mathbf{H}^k + \mathbf{U}^k \tag{5.13}$$

There are various formula for calculating $\mathbf{U}^k$, but the one suggested by Gill[72] for second derivative approximations is widely used and is known as the BFGS update:

$$\mathbf{H}^0 = \mathbf{0} \tag{5.14}$$

so that the first iteration is equivalent to the first iteration of the basic Gauss-Newton method and for the following iterations,

$$\mathbf{U}^k = \frac{1}{\alpha^k \left[\mathbf{y}^k\right]^{\mathrm{T}} \Delta\rho^k} \mathbf{y}^k \left[\mathbf{y}^k\right]^{\mathrm{T}} - \frac{1}{\left[\Delta\rho^k\right]^{\mathrm{T}} \mathbf{W}^k \Delta\rho^k} \mathbf{W}^k \Delta\rho^k \left[\Delta\rho^k\right]^{\mathrm{T}} \mathbf{W}^k \tag{5.15}$$

where

$$\mathbf{W}^k = \left[\mathbf{J}^{k+1}\right]^{\mathrm{T}} \mathbf{J}^{k+1} + \mathbf{H}^k \tag{5.16}$$

and

$$\mathbf{y}^k = \left[\mathbf{J}^{k+1}\right]^{\mathrm{T}} \left[\mathbf{f}^{k+1} - \mathbf{v}_o\right] - \left[\mathbf{J}^k\right]^{\mathrm{T}} \left[\mathbf{f}^k - \mathbf{v}_o\right] \tag{5.17}$$

The interested reader is referred to Gill[72] for further information on quasi-Newton methods applied to approximating **B** and to one of Jennings[71] or Gill[69] for further study of the BFGS and other quasi-Newton updates.

5.2.5. The Steepest Descent Method

The steepest descent method replaces the approximated Jacobian matrix in the Gauss-Newton method by the identity matrix in equation (5.9) to obtain the new direction

$$\Delta\rho^k = -\left[\mathbf{J}^k\right]^{\mathrm{T}} \left[\mathbf{f}^k - \mathbf{v}_o\right] \tag{5.18}$$

The term on the right hand side is the gradient vector, **g** and so the steepest descent method moves in the direction of the negative gradient. As the second derivatives (Hessian matrix) in the Taylor series expansion of equation (5.4) are not used, the steepest descent is a first derivative method. The term 'steepest descent' is obtained as a descent along the negative gradient maximises the reduction in the error generated by a first derivative method[66]. An optimal step length, as explained in the damped Gauss-Newton method, usually determines the amount moved in the steepest descent direction.

Progress of the steepest descent minimisation is fast far from the optimum, but slow near the optimum as a large number of steps are needed in the vicinity of the

optimum[69]. The method of steepest descent is not generally used in practice, but the Levenberg-Marquardt method discussed in the following section incorporates its best features.

5.2.6. The Levenberg-Marquardt Method

The Levenberg-Marquardt[74] method overcomes the limitations of both the Gauss-Newton and steepest descent methods and utilises the best features of each method. Yorkey[40] first suggested using it in EIT when reconstruction was attempted with noise contaminated voltage measurements. It is now used by many EIT researchers and is a popular solution method for many optimisation problems.

The Gauss-Newton theory is based on the next iteration point being in a small neighbourhood of the current point. As such, it is not very successful when the solution is far from the starting point. The advantages, however, are rapid convergence to the solution once in the neighbourhood of the solution and the ability to detect and move away from saddle points[69]. This is largely due to the second derivative approximation it provides. The steepest descent direction's progress is good when far from the solution, but its disadvantage is the lack of progress in the vicinity of the optimum. The Levenberg-Marquardt method combines the two (equations (5.9) and (5.18)) to form a new direction defined by

$$\left[\left[\mathbf{J}^k\right]^{\mathrm{T}}\mathbf{J}^k + \lambda\mathbf{I}\right]\Delta\boldsymbol{\rho}^k = -\left[\mathbf{J}^k\right]^{\mathrm{T}}\left[\mathbf{f}^k - \mathbf{v}_0\right] \tag{5.19}$$

where the Levenberg-Marquardt parameter or regularisation parameter, λ is a positive scalar constant. If λ is chosen to be large, then the unit matrix dominates the direction and so the search direction will tend towards the steepest descent. If λ is chosen to be small, then the Gauss-Newton search direction dominates. A unit step length is used as λ has effectively replaced the Gauss-Newton step length in equation (5.11).

There are various published strategies to choose λ, but essentially λ must be chosen to satisfy the convergence criteria of equation (5.12) for each iteration. Values for λ could start off large, so that the steepest descent direction dominates, and end up small, so that the Gauss-Newton direction dominates in later iterations. Various algorithms to choose λ are examined in a later section.

Many texts treat the Levenberg-Marquardt algorithm as a trust region method. The basic idea of the model trust region approach is to accept a new search direction only if the linearised model adequately reflects the behaviour of the nonlinear function. It can be shown then, for some scalar r_t related to λ, the Levenberg-Marquardt equation (5.19) is the equivalent of solving the following linearised constrained optimisation subproblem at each iteration[69]

$$\begin{aligned} &\text{minimise } \left\|\mathbf{J}\Delta\rho+\mathbf{f}-\mathbf{v}_o\right\|_2^2 \\ &\text{subject to } \left\|\Delta\rho\right\|_2 \le r_t \end{aligned} \tag{5.20}$$

This amounts to constraining the length of the new search direction to be less than r_t. In other words, the search direction is restricted to be within a ball, of radius r_t, around the current point. The new point arising from the search direction is thus only "trusted" to be within this region.

Dobson[75] examines the convergence of the Levenberg-Marquardt method in context of the inverse conductivity problem. In proving the convergence, he emphasises that convergence of the method does not imply successful reconstruction of the original distribution and as yet, nothing is known about the accuracy of the reconstruction obtained by the Levenberg-Marquardt algorithm. Reconstruction success on an EIT pressure mat is demonstrated in a later chapter, but first, another reconstruction method is examined in the following section.

5.2.7. The Scaled Levenberg-Marquardt Algorithm

Marquardt[74] recommends scaling the equations to improve the solution accuracy. This is done by normalising equation (5.19), which results in the following direction

$$\left[\left[\mathbf{J}^k\right]^T\mathbf{J}^k+\lambda\mathbf{S}\right]\Delta\rho^k=-\left[\mathbf{J}^k\right]^T\left[\mathbf{f}^k-\mathbf{v}_o\right] \tag{5.21}$$

The only change is a new matrix **S** replacing the identity matrix. Matrix **S** depends on the type of scaling used and is diagonal in structure. Marquardt[74] used the following scaling

$$\mathbf{S}_{ii}^k=\left\|\mathbf{J}_i^k\right\|_2^2 \tag{5.22}$$

where $\mathbf{J}_i$ is the ith column vector in the Jacobian and the term $\|\mathbf{J}_i^k\|_2^2$ simply equates to the ith diagonal element of $\mathbf{J}^T\mathbf{J}$. Moré[76] examined three types of scaling in his study of the Levenberg-Marquardt algorithm. The following was termed adaptive scaling and is given by

$$\begin{aligned} \mathbf{S}_{ii}^0 &= \|\mathbf{J}_i^0\|_2^2 \\ \mathbf{S}_{ii}^k &= \max\left(\mathbf{S}_{ii}^{k-1}, \|\mathbf{J}_i^k\|_2^2\right), \quad k \geq 1 \end{aligned} \tag{5.23}$$

He showed that adaptive scaling is theoretically and experimentally superior to the Marquardt scaling of equation (5.22) and others.

5.2.8. Regularised Gauss-Newton Methods

In the 1960's a number of Soviet mathematicians began research in a field of numerical analysis which used what they termed "regularisation". Regularisation methods are designed for solving ill-conditioned problems and much of the Soviet research is documented in a book by Morozov[77]. Both Hua[78] and Breckon[79] point out that the Levenberg-Marquardt method is actually a regularisation method. In this section another regularisation method published by Hua[78] is considered.

The general regularisation method reduces the negative effects that noise has on the reconstruction. The objective function being minimised is altered by adding a penalty term to offset noise induced instability so that equation (5.1) becomes

$$\Phi = \tfrac{1}{2}\|\mathbf{f} - \mathbf{v}_o\|_2^2 + \tfrac{1}{4}\lambda\|\mathbf{E}\Delta\rho\|_2^2 \tag{5.24}$$

Expanding and defining $\mathbf{S} = \mathbf{E}^T\mathbf{E}$ gives

$$\Phi = \tfrac{1}{2}\|\mathbf{f} - \mathbf{v}_o\|_2^2 + \tfrac{1}{4}\lambda\,\Delta\rho^T\mathbf{S}\Delta\rho \tag{5.25}$$

This new objective function leads to a search direction given by equation (5.21), which for a special case of $\mathbf{S}$, is the scaled Levenberg-Marquardt direction. The second term in these last two equations is the penalty term and is a quadratic function of the step direction, $\Delta\rho$. It is the objective function that is being minimised, so the role of the penalty term is to penalise large steps from the current point. The regularisation

parameter λ, controls the amount that the step direction is penalised by. Methods to select λ are considered later in Section 5.2.9.

In the general regularisation technique, **S** is any positive-definite matrix. In the unscaled Levenberg-Marquardt method, **S** is the identity matrix or a matrix of scale factors and corresponds to a penalty function of the form $\|\Delta\rho\|_2^2$. Hua[78] proposed using penalty functions obtained from forward finite differences on the elements of the finite element grid. The penalty term then either uses first differences as follows

$$\mathbf{E}\Delta\rho = \Delta\rho' \tag{5.26}$$

or second differences,

$$\mathbf{E}\Delta\rho = \Delta\rho'' \tag{5.27}$$

where each element in $\Delta\rho'$ and $\Delta\rho''$ is given by

$$\Delta\rho_j' = 2\Delta\rho_j - \Delta\rho_{j+1} - \Delta\rho_{j+a} \tag{5.28}$$

$$\Delta\rho_j'' = 4\Delta\rho_j - \Delta\rho_{j+1} - \Delta\rho_{j-1} - \Delta\rho_{j+a} - \Delta\rho_{j-a} \tag{5.29}$$

and a is the number of elements across the rectangular finite element mesh. The above equations must use the natural or row by row numbering for the finite element mesh, as shown in Figure 4.1 The matrices **E**, and consequently **S**, are constructed from equations (5.26) to (5.29) for use in calculating the regularised search direction.

The penalty term based on $\Delta\rho$ is used when the domain is known to be fluctuating, $\Delta\rho'$ is used for a continuous domain, and $\Delta\rho''$ is used when it is known to be smooth. Hua[78] showed the superiority of the latter two penalty terms over the first for one test distribution. A comprehensive study of reconstructions obtained through the use of regularisation applied to the set-up specially tailored to pressure measurement follows in Chapter 6. The regularised Gauss-Newton method is summarised in flow chart form in Figure 5.2.

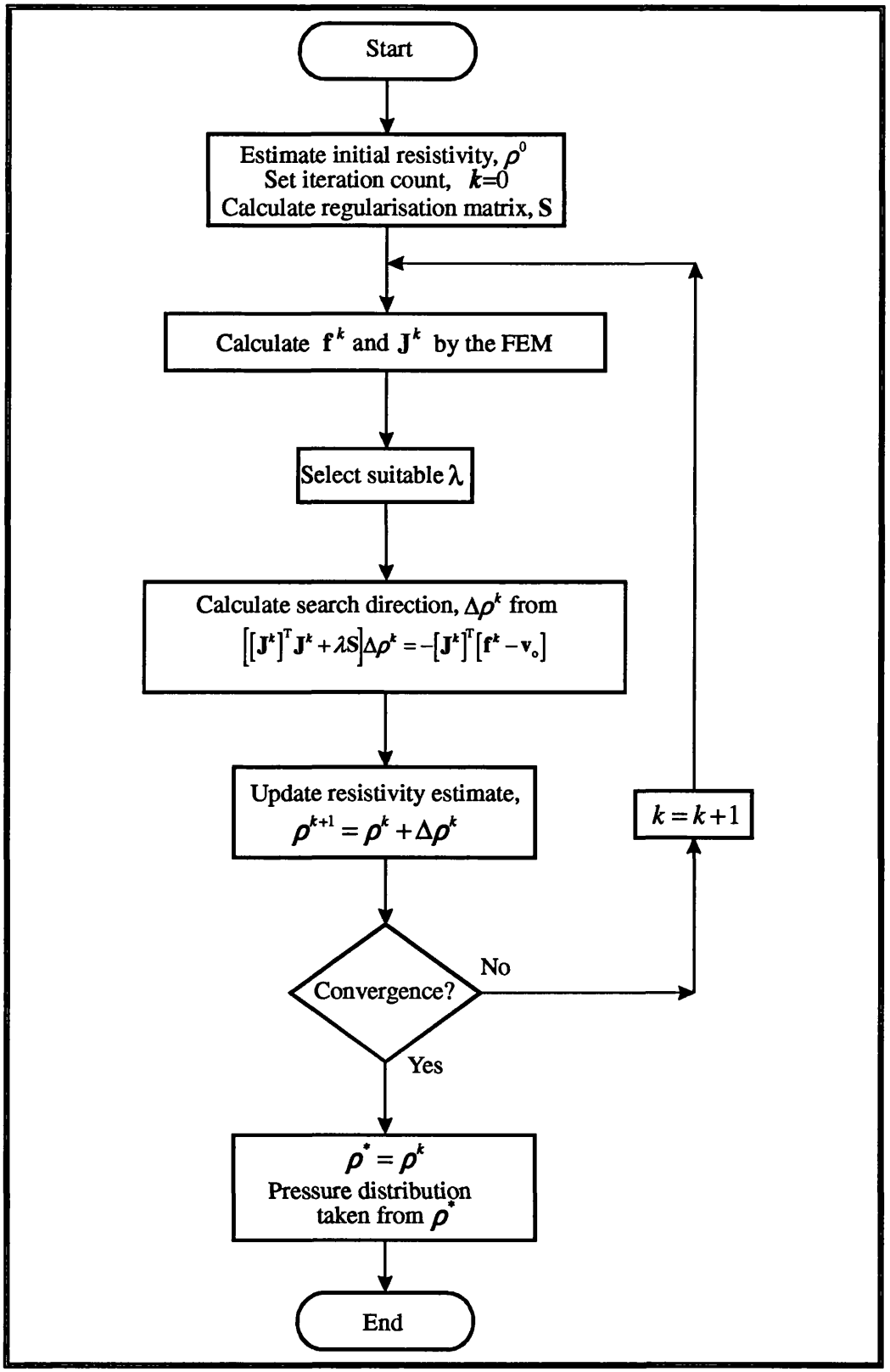


Figure 5.2 Flow Chart for the regularised Gauss-Newton method.
Note that when S=I, the procedure reduces to the Levenberg-Marquardt method.

5.2.9. Determining the Regularisation Parameter

There are a multitude of different published algorithms for finding the regularisation parameter, λ. The first was due to Levenberg[80] who used a time consuming linear search to find the optimum value for λ. Since then a few heuristic approaches have been published, but the most recent algorithms are based on a trust region approach. Since it is almost impossible to know which is best for the problem and regularisation in hand, and since it is not clear what other EIT researchers use and why; a few of these algorithms are presented and tested. The chosen algorithms include those from the simple, heuristic and trust region types.

5.2.9.1. Marquardt's Choice for the Regularisation Parameter

Marquardt[74] used a heuristic algorithm for obtaining λ, which can be stated as follows,

Choose initial λ^0, a multiplier, v_m and a divider, v_d.
Set iteration counter, k=0
  Compute $\Phi(\lambda^k)$
  If $\Phi(\lambda^k) < \Phi(\lambda^{k-1})$, let $\lambda^{k+1} = \lambda^k / v_d$ and accept ρ^k
  If $\Phi(\lambda^k) \geq \Phi(\lambda^{k-1})$, let $\lambda^{k+1} = v_m \lambda^k$ and reject ρ^k
Increment k and repeat until convergence.

5.2.9.2. Moré's Choice for the Regularisation Parameter

Moré[76] adopts a trust region approach to the Levenberg-Marquardt algorithm. Hence λ is found in terms of the trust region defined by r_t in equation (5.20). The choice of r_t at each iteration is based on a ratio calculated for a particular search direction. The ratio is between the actual reduction in the function and the reduction predicted by the linearised model. If the ratio is close to unity, that is, the two models are in close agreement, the trust region is increased. Otherwise, the trust region is decreased or kept constant, depending on the amount of agreement between the two models. Calculating the degree of change in the trust region is involved and the interested reader is referred to the paper by Moré[76]. Moré's trust region approach is highly regarded in the optimisation literature and the source code is publicly available.

5.2.9.3. Davies' Choice for the Regularisation Parameter

Davies[81] points out that when the objective function is a quadratic, the end points of the search direction follows a spiral as λ varies from zero to infinity, that is, starting at the

Gauss-Newton point and ending at the current point. Davies[81] attempts to follow this spiral and finds an estimate for λ which gives the point where the objective function is a minimum on the spiral. The algorithm is as follows:

At each iteration k, compute the gradient, $\mathbf{g}$ and the approximated Hessian, $\mathbf{J}^T\mathbf{J}$.
Then use $\lambda^k = \frac{3\mathbf{g}^T\mathbf{J}^T\mathbf{J}\mathbf{g}}{\mathbf{g}^T\mathbf{g}}$ for computing the new search direction.

5.2.9.4. The Dwindling Regularisation Parameter

The previous methods for estimating a good value for λ are involved and require many iterations to convergence. To overcome the time taken to obtain an absolute minimum, the following approach can be taken. The search direction is progressively moved from the direction of steepest descent to the Gauss-Newton direction. The iterative procedure is terminated early, that is as soon as divergence occurs. An absolute minimum is not found, but experience showed that the resistivity distribution obtained was not too dissimilar from the best possible. The algorithm is then

Choose initial λ^0 and a divider, v_d.
At each iteration, use $\lambda^k = \lambda^{k-1}/v_d$.
Terminate iterative procedure on divergence, i.e. when $\Phi(\lambda^k) > \Phi(\lambda^{k-1})$

5.2.9.5. The Constant Regularisation Parameter

Many EIT researchers seem not to vary the regularisation parameter, rather they use the following simple type of approach:

Choose λ^0 and keep constant for each iteration.
Stop on divergence or convergence, whichever is first.

5.2.10. Solution Bounds

The mat has a known resistivity with no loading. When a pressure is applied to the mat, the resistivity will only decrease in relation to the applied pressure. There is thus a maximum value for an element's resistivity. This maximum value is then used as a bounds constraint on the estimation of an element's resistivity after using equation (5.10) to update the resistivity estimation at any iteration.

5.2.11. Convergence Considerations

A decision must be taken to stop the iterative procedure. This will depend on the algorithm used to determine the regularisation parameter. For many, the procedure is stopped when improvements in the reconstructed distribution are very small. The exception is the dwindling regularisation parameter method when the procedure is terminated on divergence. Many different stopping criteria could be used and in the implemented case the iterative procedure was halted when

$$\left|\Delta\rho_i^k\right| < t \quad \text{for all } i,\ i = 1 \text{ to } m \tag{5.30}$$

where t is a tolerance level. A value of $t = 0.001$ was usually used as it gives an accuracy to 3 decimal places for conductivities in the region of unity. Any further accuracy was deemed unnecessary.

5.3. Solution of Forward Problem

To solve the forward problem, the vector **f** that will accurately estimate the voltages measured from the continuous resistivity distribution must be found. The FEM is needed to find **f**. The formation of this vector for many projection angles is discussed.

Obtaining a complete voltage measurement set requires placement of the current sources at several different electrodes. The voltages are measured at all the electrodes for each applied current pattern. Any one specific current source location or applied pattern is commonly known as a projection or projection angle. A total of p projection angles are used. Electrical Impedance Tomography is an offspring from X-ray Computed Tomography and the term 'projection angle' actually originates from this type of tomography. In X-ray Computed Tomography a projection angle is taken as the angle of the X-ray transmitter from an origin as it moves around the subject being imaged.

The current patterns being applied could consist of point injection or optimal current patterns. In point injection, a current source is connected between two electrodes and voltages are measured between sets of other electrodes. However, in this case, if the current electrodes are used for voltage measurement, any electrode contact impedance will cause a voltage drop across the electrode. Some researchers use the four electrode method to avoid the introduction of this error. This involves using two isolated points at each electrode, where one is for voltage measurement and another for current injection. Alternatively, voltage measurements on the (two) current carrying electrodes can be

avoided by taking only n readings out of the e electrodes (n=e-2). With optimal current patterns, currents are injected into all electrodes and voltages measured from all electrodes so $n = e$.

Chapter 4 examines solving the FEM equations for one projection angle. However, voltages must be solved for more than one projection angle and the following procedure is used. From the FEM, the master stiffness matrix is assembled and the following equation is formed:

$$\mathbf{YV} = \mathbf{C} \tag{5.31}$$

where $\mathbf{Y}$ is the $r \times r$ master stiffness matrix, evaluated at ρ^k, $\mathbf{V}$ is an $r \times p$ matrix, and $\mathbf{V}_i$ is a column vector containing the nodal voltages for the ith projection angle. $\mathbf{C}$ is an $r \times p$ matrix, where $\mathbf{C}_i$ contains the nodal currents for the ith projection angle. Thus $\mathbf{C}_i$ holds the natural boundary conditions. Matrices $\mathbf{Y}$ and $\mathbf{C}$ are altered by implementing the rigid boundary conditions as shown in Section 4.2.2.2.4 on matrix assembly. By enforcing these boundary conditions, $\mathbf{Y}$ becomes positive definite allowing $\mathbf{V}$ to be solved. An rp-dimensioned vector $\mathbf{v}$ is formed from $\mathbf{V}$,

$$\mathbf{v} = \mathrm{vec}(\mathbf{V}) = \left[\mathbf{V}_1^{\mathrm{T}}\mathbf{V}_2^{\mathrm{T}}\cdots\mathbf{V}_p^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.32}$$

Vector $\mathbf{v}$ represents the estimated voltage at each and every node for all projection angles. If only n electrode nodes are available for measuring voltages, the np-dimensioned vector $\mathbf{f}$, containing the estimated voltage measurements, is formed

$$\mathbf{f} = \mathit{extract}(\mathbf{v}) \tag{5.33}$$

The result of the function *extract* is the selected n rows of its argument, stacked one after another. In other words, the (r - n) unwanted rows or nodes for each projection p in $\mathbf{v}$ are removed. The resulting elements comprise a new vector $\mathbf{f}$.

The vector $\mathbf{f}$ is thus obtained via the master stiffness matrix $\mathbf{Y}$, which in turn is formed from the m resistivity elements in the FEM.

5.4. Jacobian Calculation

The Jacobian can be approximated or calculated accurately. The advantages and disadvantages of each approach needs examining in terms of time to obtain the Jacobian and the effect that an approximation has on the reconstruction algorithm.

5.4.1. Accurate Jacobian Calculation

An efficient method for accurately calculating the Jacobian matrix is presented.

The Jacobian matrix, **J**, is obtained from first derivatives by differentiating equation (5.33) with respect to ρ as equation (5.3) shows,

$$\mathbf{J} = \mathbf{f}' = \left(extract(\mathbf{v})\right)' \tag{5.34}$$

or alternatively

$$\mathbf{J} = \mathbf{f}' = extract(\mathbf{Q}) \tag{5.35}$$

where $\mathbf{Q}$ is an $rp \times m$ partitioned matrix comprising mp r-dimensioned vectors, $\mathbf{q}_{ij}$ calculated by differentiating each voltage solution vector in equation (5.31),

$$\mathbf{q}_{ij} = \frac{\partial \mathbf{Y}^{-1}}{\partial \rho_j}\mathbf{C}_i = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{Y}^{-1}\mathbf{C}_i = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{V}_i \tag{5.36}$$

There are p partitions in total, each of which corresponds to a different projection angle. Yorkey[40] demonstrated this second method of forming the Jacobian matrix and called it the standard method. Both Yorkey[40] and Woo[65] solve for vector $\mathbf{q}_{ij}$ by repeated forward and backward substitution on

$$\mathbf{Y}\mathbf{q}_{ij} = -\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{V}_i \tag{5.37}$$

All elements are solved for, but only n elements in this $r \times 1$ vector are required, so these are extracted and put into a new $n \times 1$ vector $\mathbf{j}_{ij}$,

$$\mathbf{j}_{ij} = extract\left(\mathbf{q}_{ij}\right) \tag{5.38}$$

Then all the $\mathbf{j}_{ij}$ vectors form the $np \times m$ partitioned Jacobian matrix as shown in Figure 5.3.

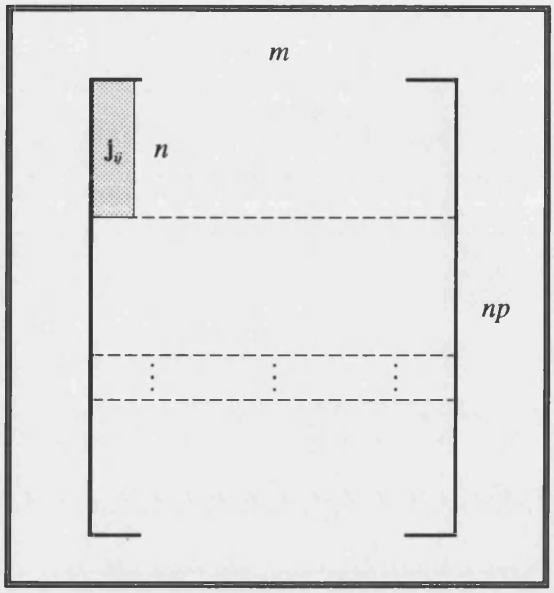


Figure 5.3 Jacobian matrix structure showing one $\mathbf{j}_{ij}$ vector

An alternative method would be to use equation (5.34) to form the $\mathbf{j}_{ij}$ vector directly, thereby avoiding the calculation of the unwanted elements in $\mathbf{q}_{ij}$ as only n rows out of r are needed in the Jacobian. Substantial savings over the previous method are possible as $n << r$ when using grids with large numbers of elements. Equation (5.36) is modified to calculate $\mathbf{j}_{ij}$ directly:

$$\mathbf{j}_{ij} = -\mathbf{G}\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{V}_i \tag{5.39}$$

where $\mathbf{G}$ is an $n \times r$ matrix and contains just the required n rows of $\mathbf{Y}^{-1}$ to form $\mathbf{j}_{ij}$. As $\mathbf{Y}^{-1}$ is a symmetric matrix, $\mathbf{G}$ can be formed from either the columns or the rows of $\mathbf{Y}^{-1}$. Columns can be obtained very efficiently through forward and backward substitution of the equation $\mathbf{Y}\mathbf{Y}^{-1} = \mathbf{I}$, where $\mathbf{I}$ is the $r \times r$ identity matrix. This enables the calculation of just the required columns in $\mathbf{Y}^{-1}$ to form $\mathbf{G}$ through selection of the appropriate columns in $\mathbf{I}$. Section 4.2.2.4.2. has already identified that sparse vector methods may also be taken advantage of during the forward substitution stage as the identity matrix consists of singleton vectors.

Matrix **G** need only be calculated once for each Gauss-Newton iteration allowing it to be re-used over and over for each vector in equation (5.39). The reason being $\mathbf{Y}^{-1}$ is the same for all projection angles when the grounded node is kept in the same position for all projection angles, see Section 6.1.1.1. The factored matrix of **Y** is obtained during the electrode voltage calculation and so may be re-used here for the forward and backward substitution stage in the solution of **G**.

The matrices on the right hand side of equation (5.39) are not explicitly formed and full matrix multiplication is not performed on them. This is because, as Yorkey[40] points out, the matrix derivative, $\partial \mathbf{Y}/\partial \rho_j$ is sparse which allows for further computational savings to be made by only performing the non-zero multiplications. The bilinear interpolation function used produces the element stiffness matrix, equation (A4.15), which has 16 elements and thus only contributes 16 non-zero elements to **Y** as well as its derivative. This is the case no matter how large the dimensions of either **Y** or $\partial \mathbf{Y}/\partial \rho_j$.

The number of floating point operations show the computational savings that can be made. The direct method based on (5.34) needs $16mp(n+1)+nz_1$ operations, whereas Yorkey's standard method based on equation (5.35) requires $mp(16+z_2)$. Each variable z_1 & z_2 contains the sum of forward and backward substitution operations in the appropriate sparse matrix/vector method. Full forward substitution must be applied to z_2, whereas singleton FF can used be for z_1. The comparisons are more tangible with an example. The example chosen is from a practical situation which was also used in the timing results of Section 4.2.2.5.1; m=400, p=16, n=32 and using the regular nested dissection (RND) ordering counts from this section, z_1=7649 and z_2=13176. The floating point operation's ratio for the direct method to the standard method is then 23:1. For grids with larger numbers of elements (larger m), this ratio increases as then m and z are the dominant counts.

5.4.2. Approximate Jacobian Calculations

Yorkey[40] also proposed the compensation theorem method which approximates the Jacobian matrix. It is more computationally efficient than the standard method when using single electrode current injection; approximately $15mn$ operation counts. It is, however, more efficient to use the standard method when using optimal current injection[47], i.e. when applying current to all electrodes. The Jacobian matrix can also be approximated by using finite differences[71]. In this approach, the value of the derivative

of $(\mathbf{f}(\boldsymbol{\rho})-\mathbf{v}_o)$ with respect to ρ_j is approximated by perturbing ρ_j by a small amount, h_j and evaluating $(\mathbf{f}(\boldsymbol{\rho})-\mathbf{v}_o)$ at this new perturbed point. The difference between the perturbed and unperturbed vectors divided by the perturbed amount gives an estimate for the appropriate entry in the Jacobian matrix. The (i,j)th element of the Jacobian is then

$$\mathbf{J}_{ij} = \frac{\mathbf{f}_i(\boldsymbol{\rho}+h_j\mathbf{e}_j)-\mathbf{f}_i(\boldsymbol{\rho})}{h_j} \tag{5.40}$$

using forward differences, where $\mathbf{e}_j$ is the jth column of the unit matrix, or

$$\mathbf{J}_{ij} = \frac{\mathbf{f}_i(\boldsymbol{\rho}+h_j\mathbf{e}_j)-\mathbf{f}_i(\boldsymbol{\rho}-h_j\mathbf{e}_j)}{2h_j} \tag{5.41}$$

using the slightly more accurate central difference approximation.

Although an approximated Jacobian reduces the rate of convergence from quadratic to linear, it may still lead to a faster solution time as in some cases it is quicker to approximate the Jacobian using finite differences[71]. The $(\mathbf{f}(\boldsymbol{\rho})-\mathbf{v}_o)$ vector is solved m times in the forward difference approach and double this using central differences. The forward difference approach then requires $m(pz_2+y)$ operations, where y is the sum of operations for the sparse factorisation and finite element assemblies. This can be compared with the number of operations given in the preceding section for accurately evaluating the Jacobian from first derivatives. It is quite clear that the finite difference approximation involves a greater number of operations and considering that the rate of convergence is not as fast, it is not worth using. However, the finite difference approximation to the Jacobian was used to verify that the Jacobian from first derivatives was being calculated correctly.

5.5. Summary

This chapter has looked at the theory required to implement the proposed pressure measurement system. Methods to solve the nonlinear problem of finding the mat resistivity distribution have been studied. Gradient methods are superior to direct search methods for the problem in hand. Gauss-Newton methods have been shown to be the best gradient method to base the reconstruction on. The three regularised Gauss-

Newton methods under consideration are based on the Levenberg-Marquardt method and the first and second finite difference penalty functions. The choice of the regularisation parameter is important and five methods were presented. Other significant considerations when using the Gauss-Newton method are scaling and the use of the no-load bounds constraint. The second derivative term may be needed for problems with a large residual at the solution. The choice of algorithms has been narrowed down, but the most suitable choice for use in a pressure scanner EIT system cannot be explicitly determined from a theoretical basis. The following chapter analyses these differing methods in order to choose an appropriate method.

The Finite Element Method is used to solve the forward problem of calculating the voltage distribution needed by the Gauss-Newton method. Both accurate and approximate Jacobian matrix calculations which are also based on the FEM were considered. An accurate Jacobian matrix calculation that is very efficient was presented. The calculation takes full advantage of sparse matrix techniques and removes all redundant matrix element evaluations. Approximate Jacobian calculations have slower convergence rates and are not as efficient as the proposed accurate calculation when using optimal currents.

6. Inverse Solution Implementation

This chapter primarily analyses the results of the Gauss-Newton methods described in the previous chapter. Details of a computer model of the mat to test the methods is described. The software required for obtaining the Gauss-Newton methods results is clarified. The results show which of the Gauss-Newton methods is the most accurate and covers further techniques for reducing the reconstruction time. Of prime importance, is the successful reconstruction in a practical environment of data measurement errors.

6.1. Computer Implementation

This section covers basic software details and computing environments used. Both algorithmic and implementational details of how the software was optimised for fast completion to reconstruction is described. One of the algorithmic methods details a new set of projection angles that minimises forward problem matrix calculations. The computer model of the pressure sensitive mat is also covered.

The reconstruction algorithms were implemented on Bath University's mainframe computer, which changed during the course of the research. Initially the mainframe was a vector machine, the Gould NP1, which was upgraded to a superscalar computer - the Sun Sparc 10/41 and subsequently to a slightly faster 20/51. All development was done under the UNIX operating system and written in Fortran 77 along with a few of the Fortran 90 extensions. The routines were portable amongst the different computer compilers.

The Numerical Algorithms Group Fortran Library[55] was used for implementing some numerical functions. These NAG library subprograms provide an up-to-date, robust and numerically stable set of algorithms. They were generally easy to use and saved a lot of unnecessary coding, especially considering an optimal approach could not always be determined until after implementation. The source code for the NAG libraries are not available. This would cause problems for porting the algorithms elsewhere, but it turned out that the chosen algorithms did not require the use of this library. One exception occurred, (a dense Cholesky solver), but a source code replacement was found in the LAPACK library on the internet (*http://www.netlib.org/lapack*). The NAG routines were also useful for study purposes, for example eigenvalue analysis. All library routines used are mentioned where applicable in the text.

Most of the methods described in the previous chapter are implemented into one program. An easy to use interface for the different test conditions has been implemented through the use of command line options and help. The same object code may be used for different sized meshes even though Fortran 77 does not support dynamic array allocation. This is possible as all finite element information is placed into configuration files by the mesh generator (Section 4.2.1), and the program is compiled with the largest matrix dimensions required. The Fortran 90 standard supports dynamic arrays, but the available compilers unfortunately did not fully implement the standard. Different projection angles and different electrode positions can be easily implemented as this information is also held in a data file. Different test pressure distributions are effortlessly set up in data files.

6.1.1. Reconstruction Time Improvements

One aim after taking over the basic Gauss-Newton algorithm from Knight[1], was to improve the time to convergence of the algorithm. Before any improvements were made, the actual time to convergence varied between two and four hours, depending on how loaded the mainframe was. This was in fact 29 minutes processor time. Waiting for these lengths of time made it difficult to test any program changes and would, in any case, be unacceptable in a clinical environment. The time that was needed for 5 iterations on a 25 element FEM mesh. Larger meshes resulted in impractical times to reconstruction. The initial aim was to decrease the reconstruction time and then to increase the size of the FEM mesh to a more realistic one. The times have been decreased from 29 minutes to 1 second on the Gould computer for the 25 element mesh. The Sun 20/51 proved to be faster and this example took less than a twentieth of a second. A large contributing factor for the improvement is the implementation of the sparse techniques already covered for the forward problem. The gain in speed to reconstruction is greater for meshes with larger numbers of elements. Other reasons for this gain in performance are briefly discussed.

The most dramatic computer optimisations are due to algorithmic improvements. However, smaller increases in efficiency are also due to the following sensible implementational techniques.

- Minimise magnetic file accesses by reading data from disk just once then storing in RAM, which is now quite abundant in modern computers.

- Use single arrays for n-dimensioned vectors, rather than $n \times 1$ matrices.
- Expanding subroutines embedded within large time critical loops into inline code, thereby removing subroutine call overhead.
- Contiguous memory accesses when referencing many matrix elements.
- Storing variables in memory thus avoiding re-calculation.
- Implementing individual matrix multiplication routines that take advantage of any special structure. For example, the matrix $\mathbf{J}^T$ need not be calculated from $\mathbf{J}$ for the multiplication $\mathbf{J}^T\mathbf{J}$ and as this matrix is symmetric, just the upper triangle need be calculated via a Cholesky solver. Even after all optimisations, this matrix multiplication is the most expensive task in the regularised Gauss-Newton method.

6.1.1.1. Projection Angles

Most of the reconstruction time was spent in calculating $\mathbf{Y}$ and its inverse, so is an obvious target for optimisation. A major decrease in reconstruction time was due to the implementation of a new set of boundary conditions.

Two examples of the original natural boundary conditions are shown in Figure 6.1. In this configuration, the current is injected into one node and extracted at another earthed node. As can be seen, each projection angle has a different earth node. The voltage measurements are made between all non-current carrying electrodes and earth. In the new configuration, shown in Figure 6.2, the same node is earthed for each projection angle. Current is still injected into and extracted from the same nodes. No current flows down the earthed node as the same value of current injected into the mat can be extracted from it. This is done by floating the current source. One node must be earthed to satisfy the rigid boundary conditions. Voltage measurements are still taken between all non-current carrying electrodes and earth.

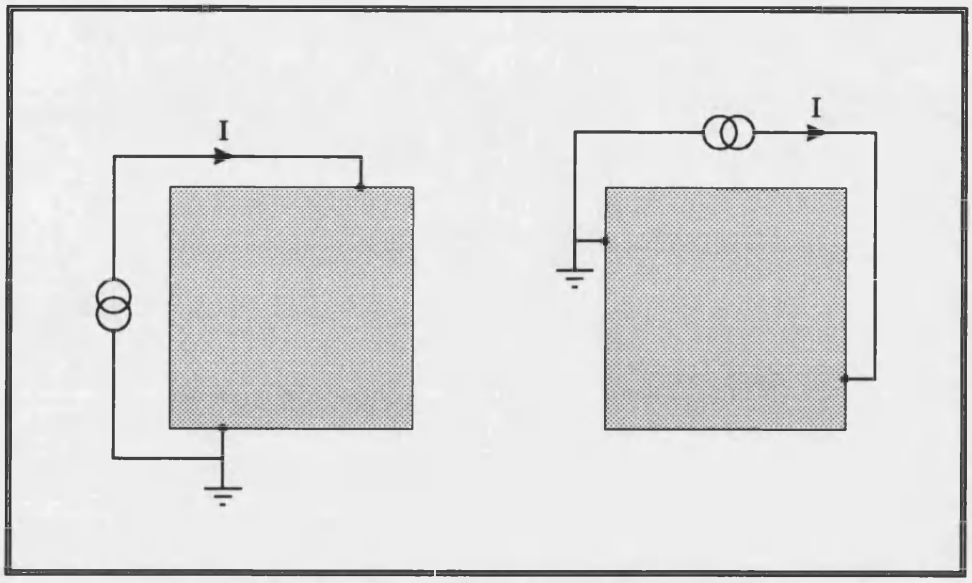


Figure 6.1 Two different projection angles

The reason for using this new rigid boundary condition, is to keep the $\mathbf{Y}^{-1}$ matrix the same for each projection angle. This allows $\mathbf{Y}^{-1}$, which is used many times over, to be calculated just once for each iteration. In practice, $\mathbf{Y}$ is factored once and forward and backward substitution is performed for each different projection angle, as described in Section 5.4.1. After implementing this single fixed rigid boundary condition and the sparse matrix techniques, time spent in the forward problem changed from a major to a minor part of the reconstruction process.

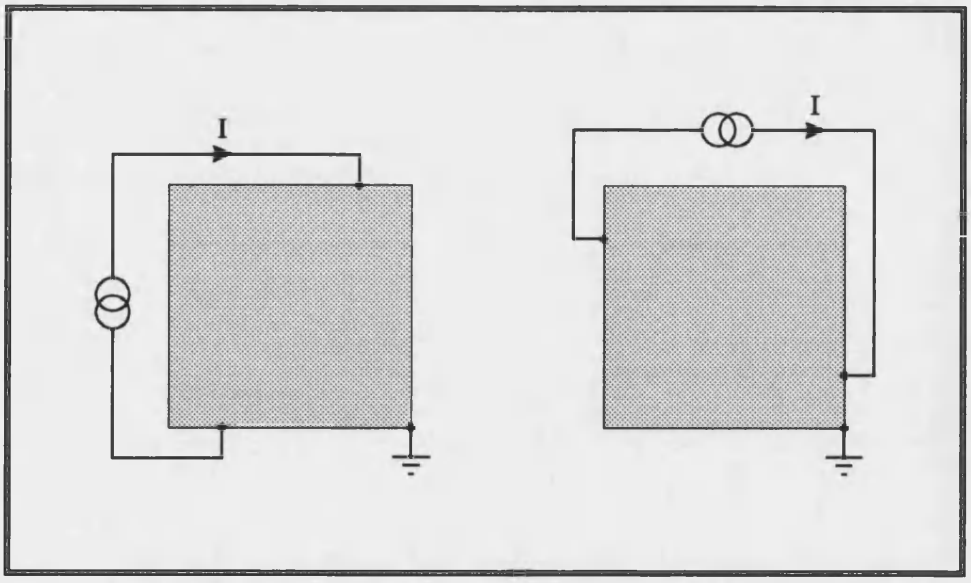


Figure 6.2 Two projection angles with one earth connection and floating current source

6.1.2. Modelling the Pressure Sensitive Mat

The pressure sensitive mat is modelled through a computer simulation. The use of a simulation rather than actual readings in EIT is common practice and allows voltage 'measurements' to be repeated, enabling proper analysis and development of the algorithms. The objective of the research work being undertaken is to establish the feasibility of using Electrical Impedance Tomography to measure interface pressure. The controlled modelling of the mat facilitates this objective.

The finite element method using bilinear square elements is the basis of the model, as discussed in Section 4.1.2. The mat's resistivity is modelled and it is a simple matter to translate the reconstructed resistivity distribution into the applied pressure distribution. A square mat is modelled as this is the shape required for wheelchair patient pressure measurement. Extension to a rectangular shape for bed-ridden patients is a trivial task, as square elements can still be used. Any shaped domain can be modelled with a combination of triangular and square elements if required.

The role of the model is purely to simulate realistic voltage measurement readings for any given applied current pattern and test pressure distribution. The computer code for this is very similar to the forward problem, which also obtains voltage readings for given current patterns and resistivity distributions. The modelling program is separate to the reconstruction program, but uses the same mesh configuration files. As with the reconstruction program, different set-ups are easily configured through command line parameters and data files produced by the mesh generator. The FEM produces exact voltages for any sized mesh model, but in a practical system these voltage measurements will be contaminated with noise. Noise was simulated by adding an error to each electrode voltage produced by the FEM, $\mathbf{v}_{FEM}$, to give a simulated voltage measurement with error, $\mathbf{v}_E$

$$\mathbf{v}_E = \mathbf{v}_{FEM}(1 + r_u E) \tag{6.1}$$

where r_u is a random number between -1 and +1 with uniform probability density function and E is the error to be simulated as a percentage.

6.2. Methods to Solve the Least Squares Equations

Every iteration requires a set of linear equations to be solved, for example, equation (5.9) for the Gauss-Newton direction. Two aspects of solving the linear equations are significant. Firstly, the speed with which they can be solved and secondly, the accuracy to which they can be solved. Speed is important as the solution of these equations plays a large part in the iterative process. A comprehension of matrix conditioning and eigenvalues is essential to the concepts presented in this thesis and is needed in this section covering equation solution. They are briefly covered in Appendix A2.

6.2.1. Cholesky Decomposition

The linear equations require the 'inversion' of a positive definite symmetric matrix. This matrix structure can be taken advantage of by factorising $\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}+\lambda\mathbf{S}\right)$ using a Cholesky decomposition followed by forward and backward substitution to solve for $\Delta\boldsymbol{\rho}^{k}$. The LAPACK routine *dposv* is a Cholesky solver and was used to solve the linear equations at each iteration. The iterative refinement it contained was removed as iterative refinement to machine precision is an unnecessary waste of time in a process that is already iterative, especially where the final solution has relatively large errors.

6.2.2. QR Decomposition

Another commonly used method to solve linear equations is through the QR decomposition, whereby a matrix is factored into the product of an orthonormal matrix, **Q** and an upper triangular matrix, **R**. For application to the least squares problem, the following augmented matrix is required

$$\mathbf{A}=\begin{bmatrix}\mathbf{J}\\ \lambda^{\frac{1}{2}}\mathbf{E}\end{bmatrix} \tag{6.2}$$

so that the regularised Gauss-Newton direction of equation (5.21) can be rewritten and simplified,

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\Delta\boldsymbol{\rho}^{k}=\mathbf{R}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{R}\Delta\boldsymbol{\rho}^{k}=\mathbf{R}^{\mathrm{T}}\mathbf{R}\Delta\boldsymbol{\rho}^{k}=-\mathbf{J}^{\mathrm{T}}\left[\mathbf{f}-\mathbf{v}_{\mathrm{o}}\right] \tag{6.3}$$

Forward and backward substitution will then solve for $\Delta\boldsymbol{\rho}^{k}$. The QR decomposition is usually recommended for solving least squares problems and many implementations use it, for example Moré[76]. It is favoured in order to avoid the matrix multiplication $\mathbf{J}^{\mathrm{T}}\mathbf{J}$, thereby eluding possible computer underflow and overflow on ill-conditioned problems.

These effects are most noticeable on highly ill-conditioned problems since, if the condition number of $\mathbf{J}$ is κ, the condition number of $\mathbf{J}^T\mathbf{J}$ is κ^2. This is particularly relevant to the unregularised Gauss-Newton equations where the condition number of $\mathbf{J}^T\mathbf{J}$ can be enormous for meshes with large numbers of elements. The Cholesky decomposition requires a non-singular matrix, so it breaks down in this case as $\mathbf{J}^T\mathbf{J}$ is singular to machine precision (see Appendix A2). When $\mathbf{J}$ is singular to machine precision, the singular value decomposition (SVD) can instead be used to solve the equations; see Gill[72] for more information on the SVD.

Regularisation overcomes the singularities that are present in badly conditioned inverse problems. The identity matrix that the Levenberg-Marquardt algorithm adds to $\mathbf{J}^T\mathbf{J}$ increases its eigenvalues by λ. An ill-conditioned problem has relatively small eigenvalues and so with a relatively large λ, regularisation can alter a set of equations from ill-conditioned to well conditioned. The final conditioning is highly dependent on the size of λ for all types of regularisation. Experience showed that to obtain a direction of descent with errors in the measured voltages, λ had to be large relative to the machine precision. A large λ gives rise to well conditioned equations and any inaccuracies, due to rounding when using the Cholesky decomposition, were not noticeable. The QR approach requires larger equations to be solved. Compared to the QR decomposition, the Cholesky decomposition takes approximately a quarter the number of floating point operations[82] and thus is the preferred method for solving the regularised equations.

6.2.3. Hachtel's Augmentation

Woo[83] suggested using what is known as Hachtel's augmented matrix method to solve the linear equations for a circular EIT resistivity model. Equation (5.21), sometimes termed the normal equation, can be simply re-written with the following augmented matrices,

$$\begin{bmatrix} \alpha\mathbf{I} & \mathbf{J} \\ \mathbf{J}^T & -\nu\mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \Delta\rho^k \end{bmatrix} = \begin{bmatrix} -[\mathbf{f}-\mathbf{v}_o] \\ \mathbf{0} \end{bmatrix} \tag{6.4}$$

where $\lambda = \alpha\nu$. The variable α cancels out when solving this set of equations, but its choice affects the conditioning of the augmented equations to be solved. A near optimal value for α is the smallest singular value of $\mathbf{J}$[84]. The augmented matrices were formed and solved by a NAG[55] symmetric indefinite equation solver. When α is set to 1 the

conditioning was found to be similar to that of $\mathbf{J}^T\mathbf{J}$. With the near optimal value for α, the augmented matrix conditioning was closer to that of the Jacobian matrix. Hachtel's method also avoids the explicit matrix multiplication $\mathbf{J}^T\mathbf{J}$ and so the same accuracy arguments for using the QR decomposition can be applied. The regularised Gauss-Newton method produces well-conditioned matrices and the results are the same, to a large number of decimal places, (see Appendix A2), when comparing with the Cholesky solution. However, when solving the unregularised Gauss-Newton equations, the Hachtel results were accurate to a larger number of decimal places, depending on the conditioning. Woo[83] found that a much more accurate reconstruction arose as a result of using Hachtel's method as opposed to the normal equations. The results do not correlate as Woo also used regularisation, but his more accurate results might be due to a highly ill-conditioned **J**, less regularisation being applied, and floating point arithmetic performed with lower precision.

Woo[83] also found Hachtel's method more suitable and faster than Gaussian elimination for implementation on a parallel system. The augmented matrix is larger but sparse due to the diagonal matrix **I** and the band matrix **S**. Hachtel's method, with an appropriate sparse matrix solver, may be faster than the Cholesky method when using a scalar processor.

6.3. Reconstruction Results

This section examines the quality of the reconstructions given by the Gauss-Newton methods. The test conditions attempt to use pressure profiles and conditions that would be found in a real-world medical pressure scanner environment.

6.3.1. Test Configurations

To determine the best reconstruction approach, a number of test distributions were tried. The test distributions used for analysing the reconstruction algorithms are presented. Methods by which error in the reconstruction can be measured are described, but firstly, the type of projection angles used are discussed.

The mat being modelled is square shaped and reconstruction of the pressure profile is ultimately reduced to finding the pressure or resistivity for each square finite element. The number of elements in the finite element grid determines the resolution in the reconstruction. The resolution required will depend on the application. Various meshes ranging from coarse to fine were considered, i.e., square meshes with 64, 100, 256 or

400 elements. The electrodes for the 64 and 256 element grids were placed in the same locations, on common grid nodes. Likewise for the 100 and 400 element meshes. Thirty-two electrodes were used for each configuration, as shown in Figures 6.3 and 6.4.

The projection angles use point current injection patterns at the electrodes. Four projection angles were used for the 64 element grids and eight projection angles for the 100 element grids. The larger 256 and 400 element grids used sixteen projection angles. This provided enough unique information for reconstruction except for the 400 element grid. Although further projection angles could be used in the latter case, it is interesting to see how the algorithms perform under these conditions.

The test profiles or test pressure distributions used in this chapter are called BLOCK, BUTTOCKS, FOOT and WHEELCHAIR and are shown in Figures 6.5 to Figure 6.8. The latter three tests are directly applicable to medical pressure measurement. Test BLOCK is included to see how well the algorithms perform on a pressure profile with a non-smooth pressure distribution.

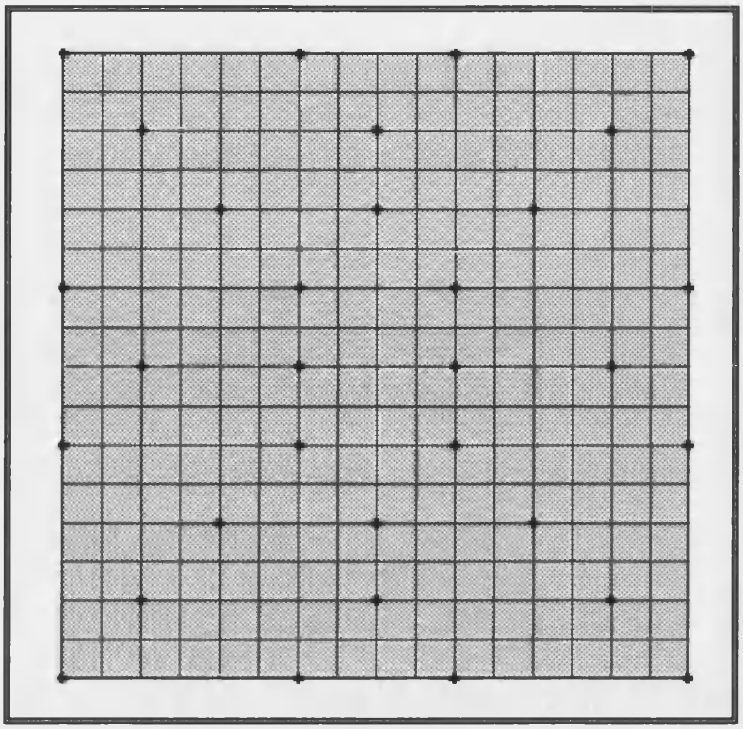


Figure 6.3 The 32 electrode configuration for grids with either 64 or 256 elements

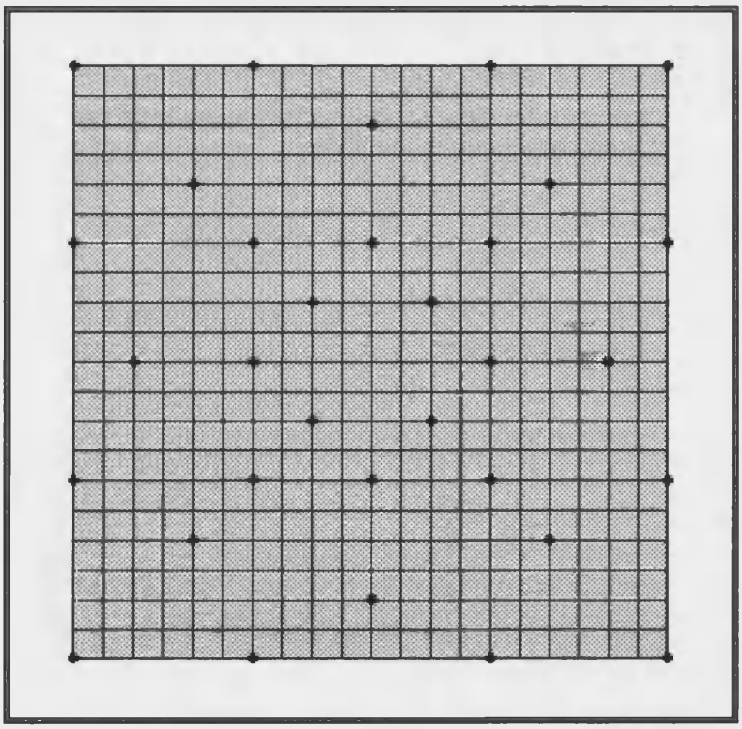


Figure 6.4 The 32 electrode configuration for grids with either 100 or 400 elements

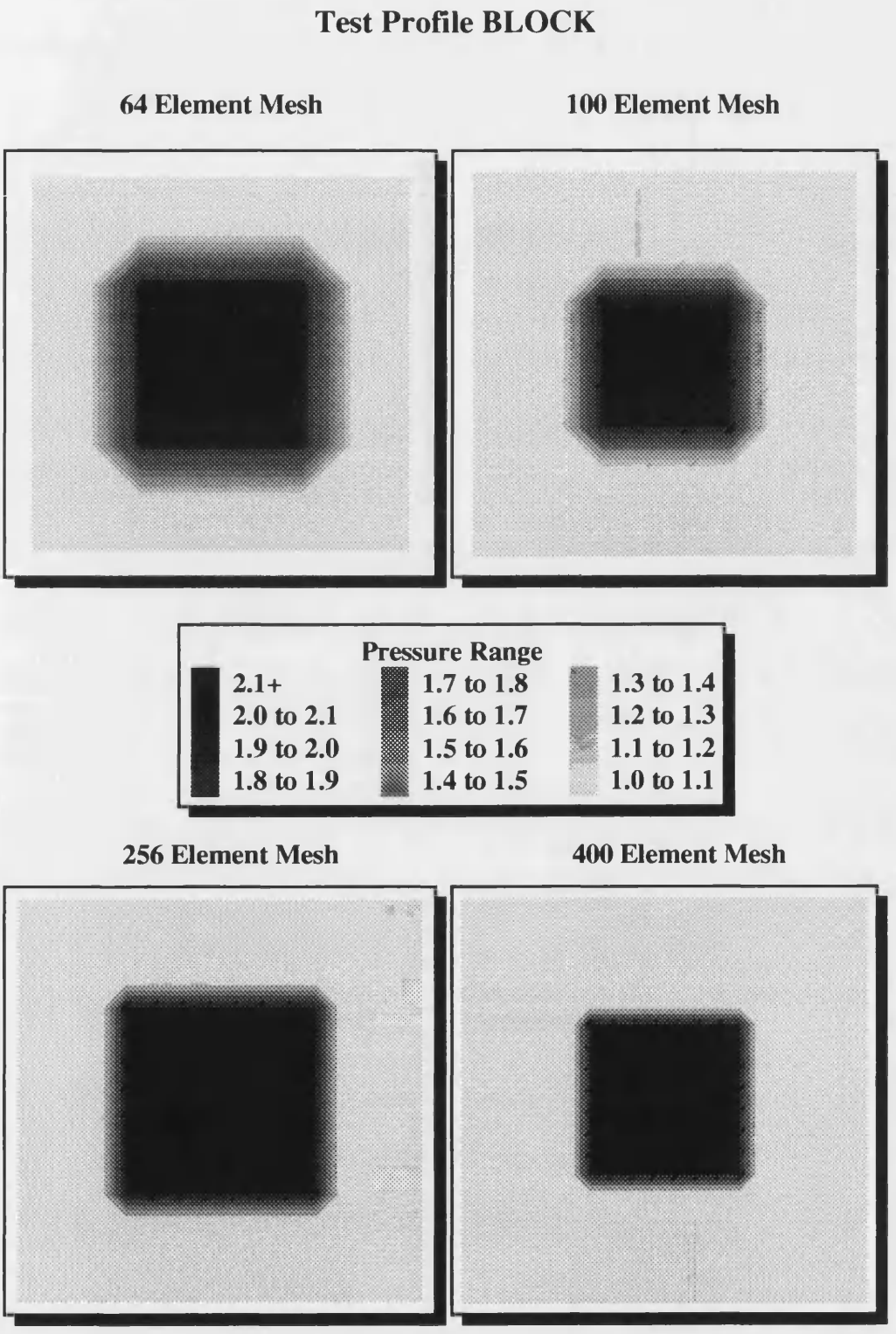


Figure 6.5 Test profile BLOCK on 4 different meshes.

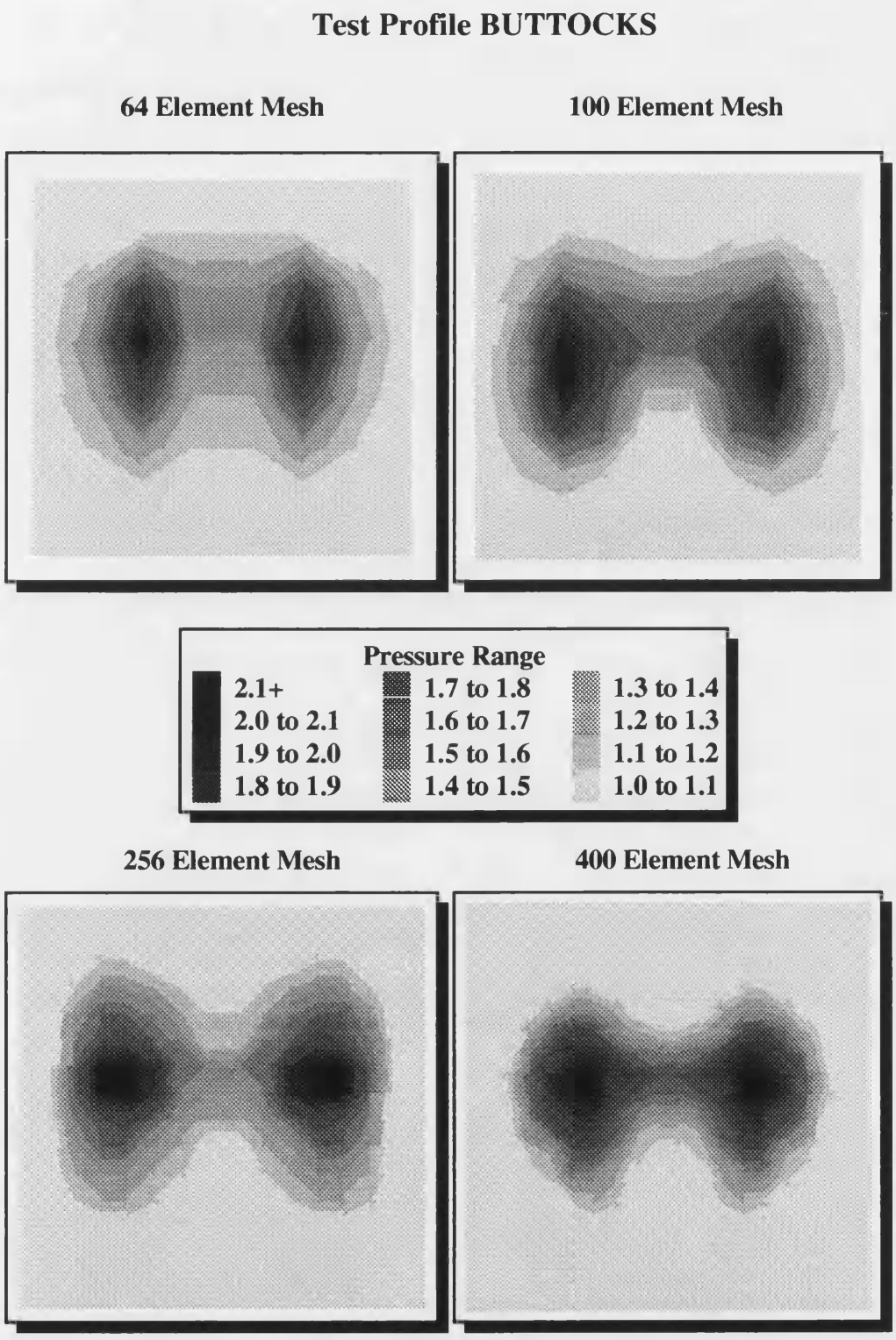


Figure 6.6 Test profile BUTTOCKS on 4 different meshes.

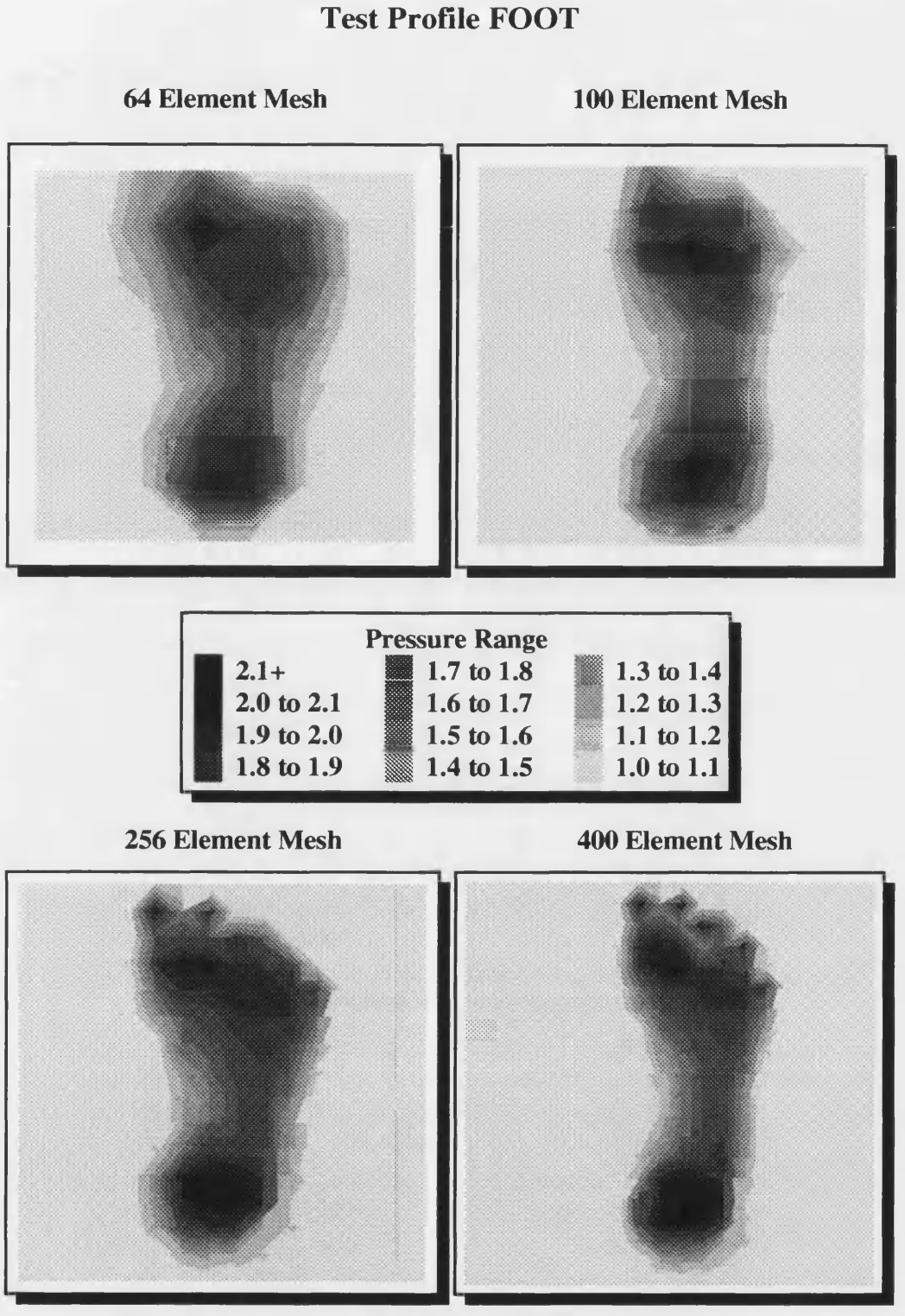


Figure 6.7 Test profile FOOT on 4 different meshes.

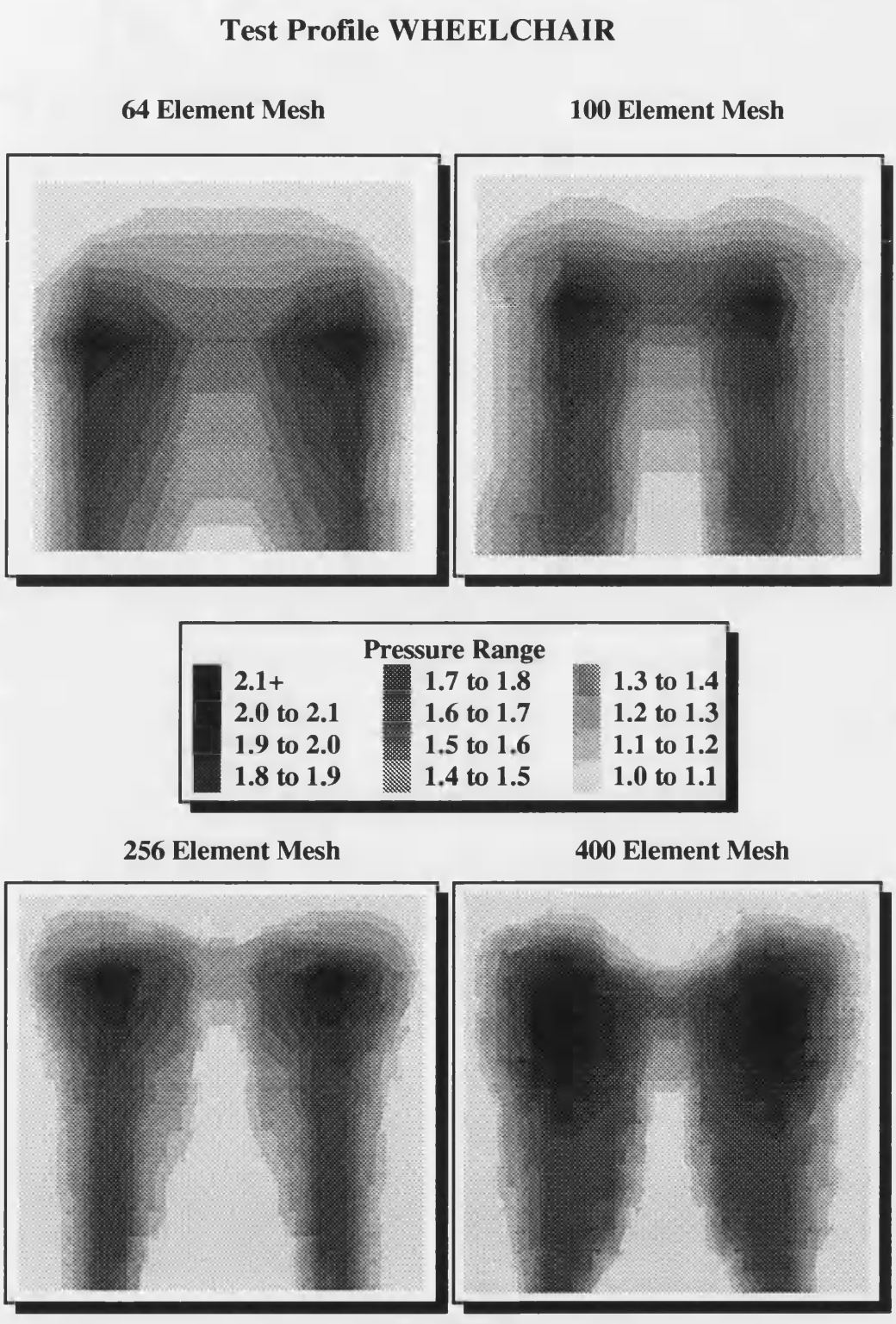


Figure 6.8 Test profile WHEELCHAIR on 4 different meshes.

The initial estimate of the pressure distribution given to the reconstruction algorithms is the uniform 'no-load' pressure. This is a reasonable starting point as the no-load pressure/resistivity will always be accurately known.

The quality of the reconstructed pressure distribution can be measured through error analysis. The element error is defined as the average error in each element compared to the true distribution. Expressed as a percentage, the element error is then

$$\text{element error} = \frac{100\sum_{i=1}^{m}\frac{abs(\rho_i^* - \rho_i^k)}{\rho_i^*}}{m} \tag{6.5}$$

where m is the total number of elements in the FEM mesh and $abs()$ is the absolute value, ρ^* is the true resistivity distribution and ρ^k is the resistivity at iteration k.

The normalised element error is the element error normalised to the initial estimate,

$$\text{normalised element error} = 100\frac{\sum_{i=1}^{m}\frac{abs(\rho_i^* - \rho_i^k)}{\rho_i^*}}{\sum_{i=1}^{m}\frac{abs(\rho_i^* - \rho_i^0)}{\rho_i^*}} \tag{6.6}$$

where ρ^0 is the initial resistivity estimate. The normalised element error is used so that all test distributions can be compared together as all have a normalised element error of 100% before the iteration starts.

Realistically quantifying the error in the image is difficult. Yorkey[47] used a two-norm error measurement to evaluate image error. However, the two-norm error does not always correlate to visible image error. The element error calculation overcomes this to some degree, but as it is an averaging function, large errors in single elements are not very noticeable. This highlights a problem with using error measurement values to judge reconstruction success. The element error value does not always correlate with the quality of the reconstruction, but it was found to be better than a two-norm error. The element error is used throughout this chapter for comparisons between various methods. The quality of the reconstructions generally relate well to the element error and the exceptions are noted where necessary.

6.3.2. Gauss-Newton and Damped Gauss-Newton Method Reconstruction Results

The Gauss-Newton method was coded using the optimal programming techniques described so far. A NAG[55] library routine, *e04fdf*, was used to study the damped Gauss-Newton reconstructions. This implementation uses a search method, described by Gill[72], to obtain a step length which approximates the minimum along the Gauss-Newton direction. Although the NAG version was much slower per iteration when compared to the undamped version specially coded for the EIT pressure problem, it was a fast track route to test the usefulness of damping. The NAG routine did not have any options for implementing solution bounds and so this was added during the forward problem calculation function call, which has to be coded by the user.

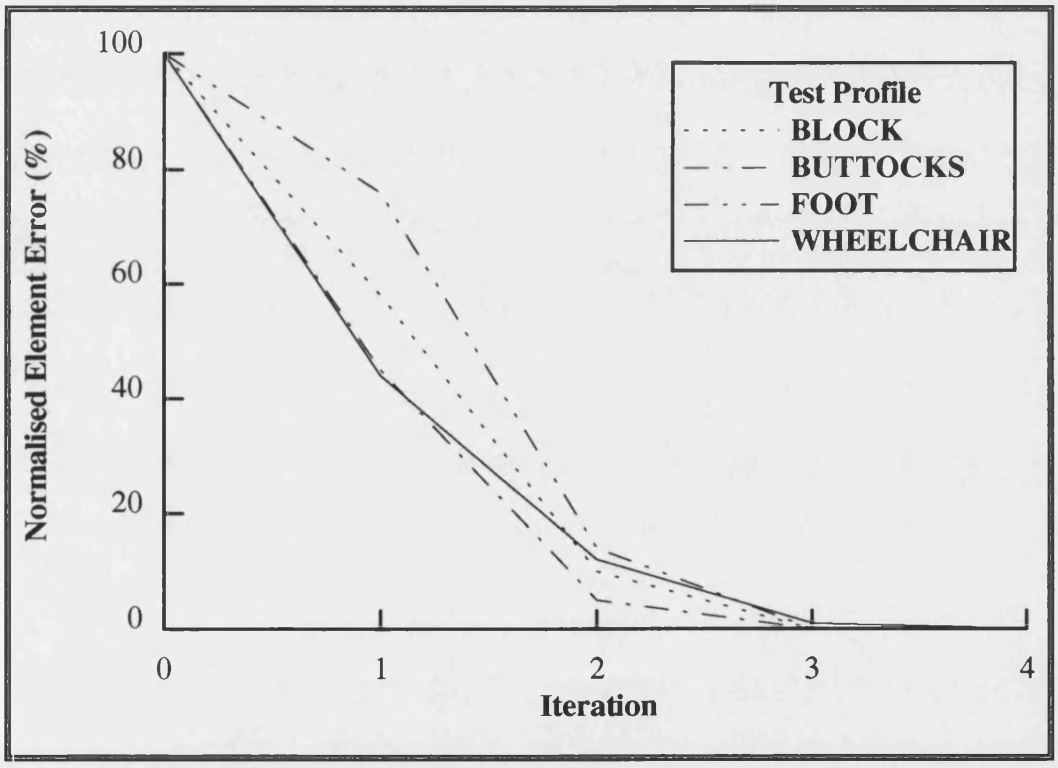


Figure 6.9 Gauss-Newton reconstruction errors on 4 different test profiles using error free voltage readings and the 100 element grid.

Monitoring the normalised element error at each iteration indicates the success of the reconstruction. This can be seen by the zero reconstruction errors at the final iteration, as shown in Figure 6.9 for the four test profiles using the 100 element mesh. Although the test profiles were perfectly reconstructed, this is in the ideal situation of zero error in

the voltage readings. The real test is when a voltage error is present in the measured signals. Figure 6.10 shows the error in a reconstruction with a range of added voltage measurement errors. The element error is that on convergence. With the higher voltage errors, the reconstruction begins to diverge after initially converging. The iterative process is halted at this point and the element error shown is then that before divergence occurs. This graph shows the rapid degradation in the reconstruction as the measurement error is increased.

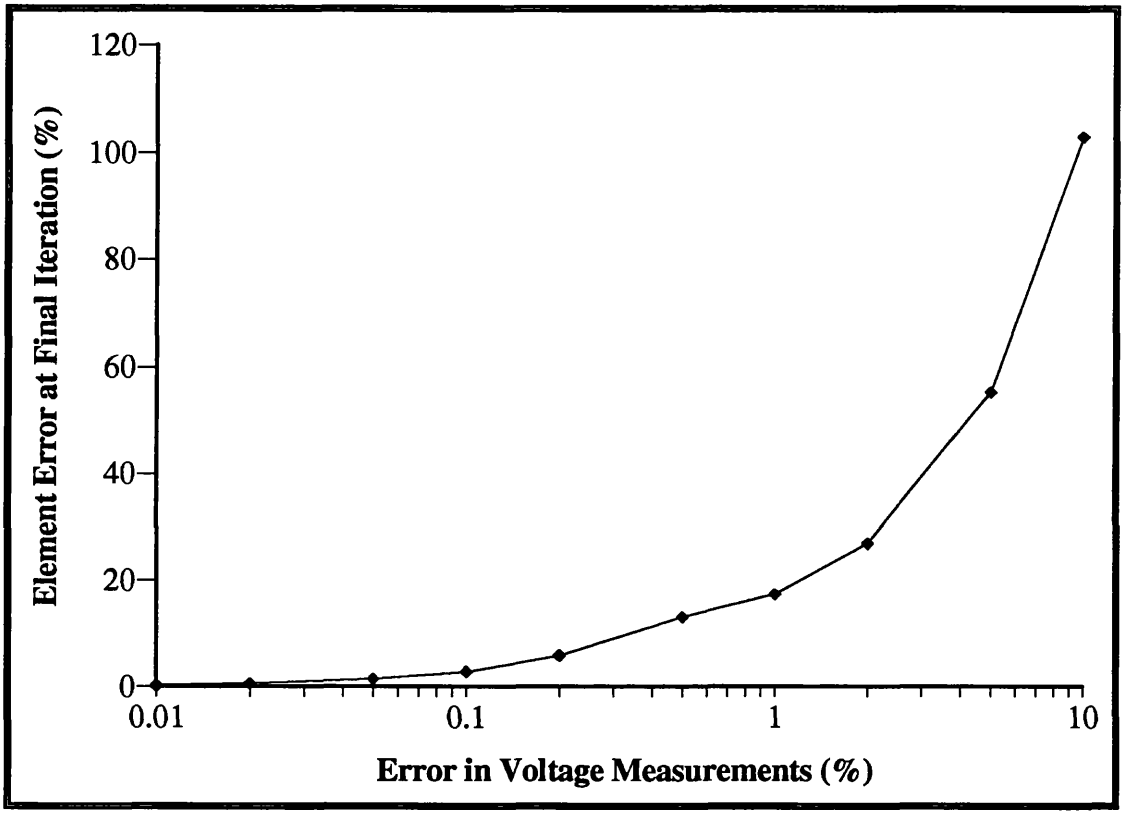


Figure 6.10 Graph of the average error in each element at the final iteration using the Gauss-Newton method for a range of added voltage measurement errors. Results are for test profile WHEELCHAIR on the 100 element mesh.

Figure 6.11 shows the final reconstruction of test WHEELCHAIR on the 100 element mesh. The reconstruction with 1% voltage error is shown as this is the amount of expected error in a practical system. As Figure 6.10 shows, the final reconstruction has an average element error of 17%. Although the average element error is not particularly high, the reconstruction clearly demonstrates that the Gauss-Newton method is of no use in a practical environment. The reconstruction bears no resemblance to the actual

distribution and the same was found for all test pressure distributions tried. This reconstruction is on the 100 element mesh. Reconstructions were either worse or failed to reconstruct on meshes with larger number of elements.

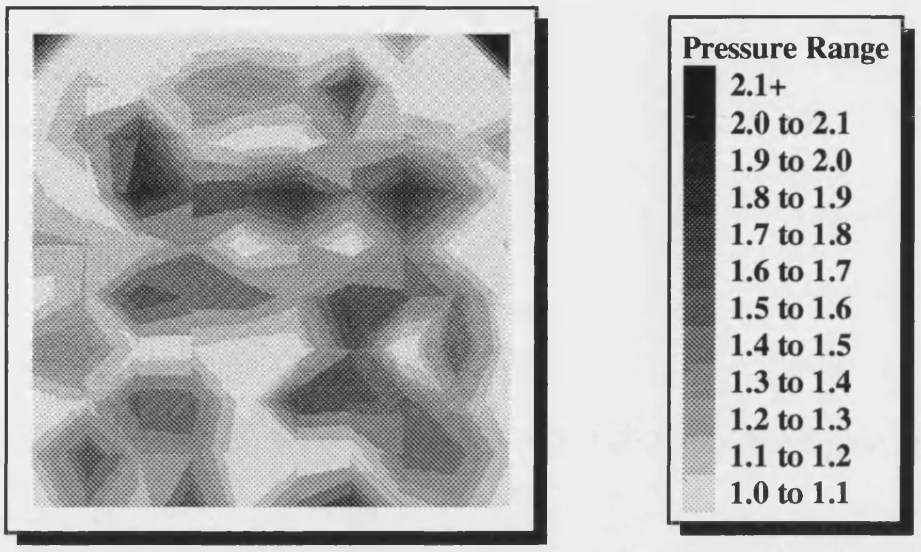


Figure 6.11 Reconstruction of the 100 element test WHEELCHAIR by the Gauss-Newton method with 1% error in the voltage readings.

Yorkey[40] first demonstrated the success of the Gauss-Newton method over others in EIT. The Gauss-Newton reconstruction is very disappointing in the presence of noise. This is well known in the EIT field and is why researchers have been trying to improve the Gauss-Newton method. The methods considered in this thesis use many of the improvements, but also considers other methods not mentioned in the literature. One of which is the effect of damping on the Gauss-Newton method, which is now examined.

Figure 6.12 is a typical example of the improvements that damping has when the Gauss-Newton method begins diverging. The use of a step size less than unity restrains the reconstruction so that divergence doesn't occur. As the damped Gauss-Newton method uses a search to find the optimum step length, the results are better, but all it effectively does is prevent divergence. Little or no improvement occurred if the damped Gauss-Newton method continued for many more iterations than the 5 shown in Figure 6.12. The next possible improvement studied is the inclusion of the second derivative term.

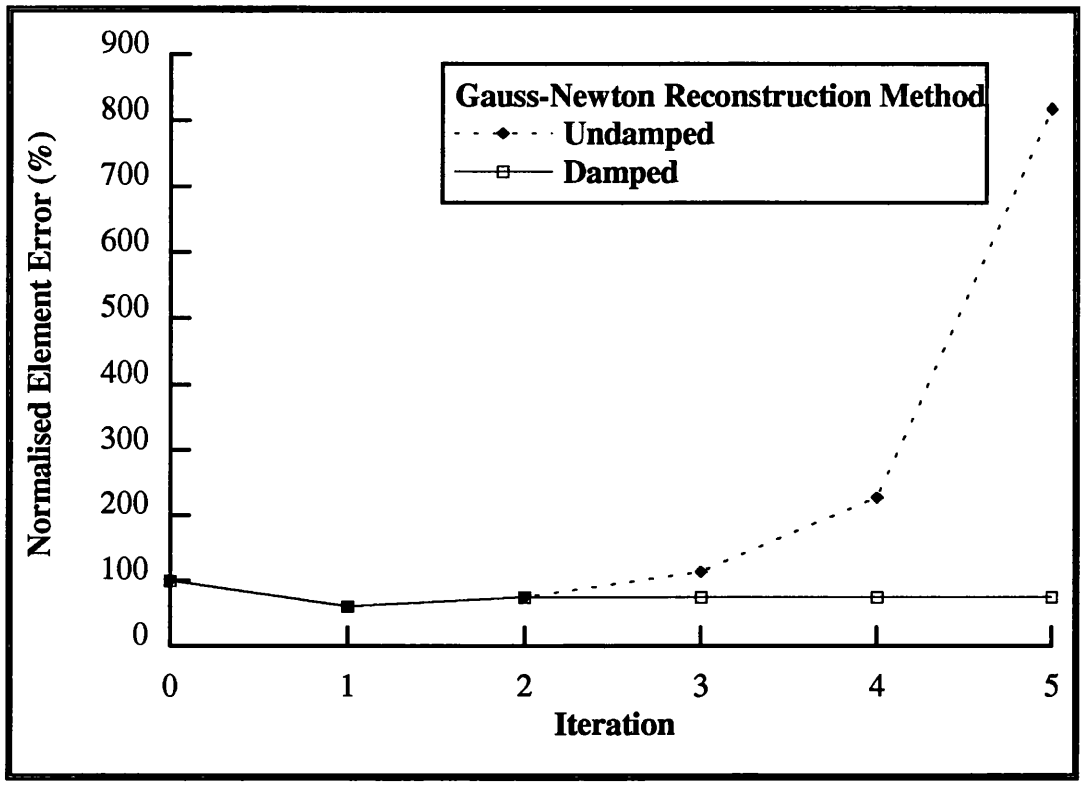


Figure 6.12 Reconstruction errors comparing damped and undamped Gauss-Newton reconstructions. The test profile is WHEELCHAIR with 0.5% added voltage error on the 100 element grid.

6.3.3. Reconstruction Results by the Gauss-Newton Method with Second Derivative Approximations

The BFGS approximation to the second derivatives were programmed into the Gauss-Newton Fortran program. The results using second derivative approximations were very disappointing. The Gauss-Newton method diverged resulting in poor or unrecognisable reconstruction.

The conditions for "large" residuals, as defined by Gill[72], occur in the presence of noise. The greater the noise, the larger the residuals will be at the solution. The principal assumption of the Gauss-Newton method is that the first derivatives accurately approximate the Hessian. This is valid with error free data, but when there is noise in the voltage measurements the first derivatives do not fully approximate the Hessian matrix. The optimisation literature urges the use of second derivative approximations and this is one area which has not been covered by the EIT literature. The disappointing results indicate that the approximation to **B** is no good. Gill[72] indicates that the BFGS

approximation to **B** can be poor. An accurate calculation of the **B** matrix would be an extremely CPU intensive task and the regularisation method in the following section are fast and more promising.

6.3.4. Reconstruction Results using Regularisation

Regularisation improves the reconstruction quite considerably. Yorkey[40] first showed the Levenberg-Marquardt regularisation improvements applied to EIT. Hua[78] showed that the finite difference type of regularised Gauss-Newton algorithm can be applied to EIT. As most EIT research is geared towards body tissue measurements or industrial process tomography, the domain being modelled is inevitably irregularly shaped, where the use of finite difference regularisation is more difficult. To the author's knowledge, this regularisation is not used in any practical EIT systems and no extensive study of different regularisation methods has been published. The aim of this section is to compare the Levenberg-Marquardt regularisation with the first and second difference regularisation methods. Results on different sized meshes are examined and tested with the particular type of electrode scheme used, i.e. with the use of internal electrode nodes, detailed in the next chapter.

The results concentrate on voltage measurements with a conservative measurement error of 1%. Hu[41] suggested that an electrode positioning error of less than 3% led to an average measured voltage error of less than 1%. The theoretical data he obtained was from the FEM, but no detail on the number of elements or type of elements was given. It is, however, an indication of the practical system errors, and a 3% electrode positioning error must be a maximum figure for the proposed resistive mat. The electrode positioning error is the greatest source of measurement error in most EIT systems.

All regularisation methods were coded into the Fortran Gauss-Newton program and use the Marquardt choice for the regularisation parameter. This choice of regularisation parameter is detailed in the following section. Figures 6.13 to Figure 6.16 show results using the three regularisation methods on four different grid sizes for the four different test profiles of Figures 6.5 to Figure 6.8. The success of the reconstruction is indicated by the normalised element error on the final iteration. The reconstructed distributions can be seen in Figures 6.17 to 6.20.

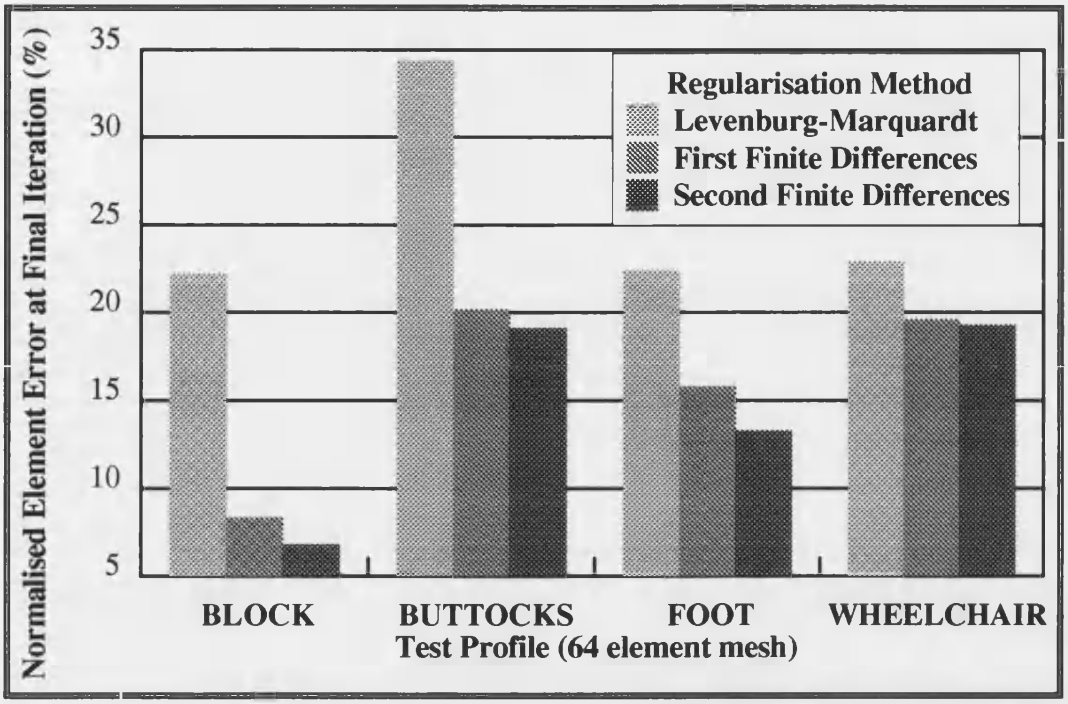


Figure 6.13 Regularisation method comparisons on the 64 element mesh with added voltage measurement error of 1%.

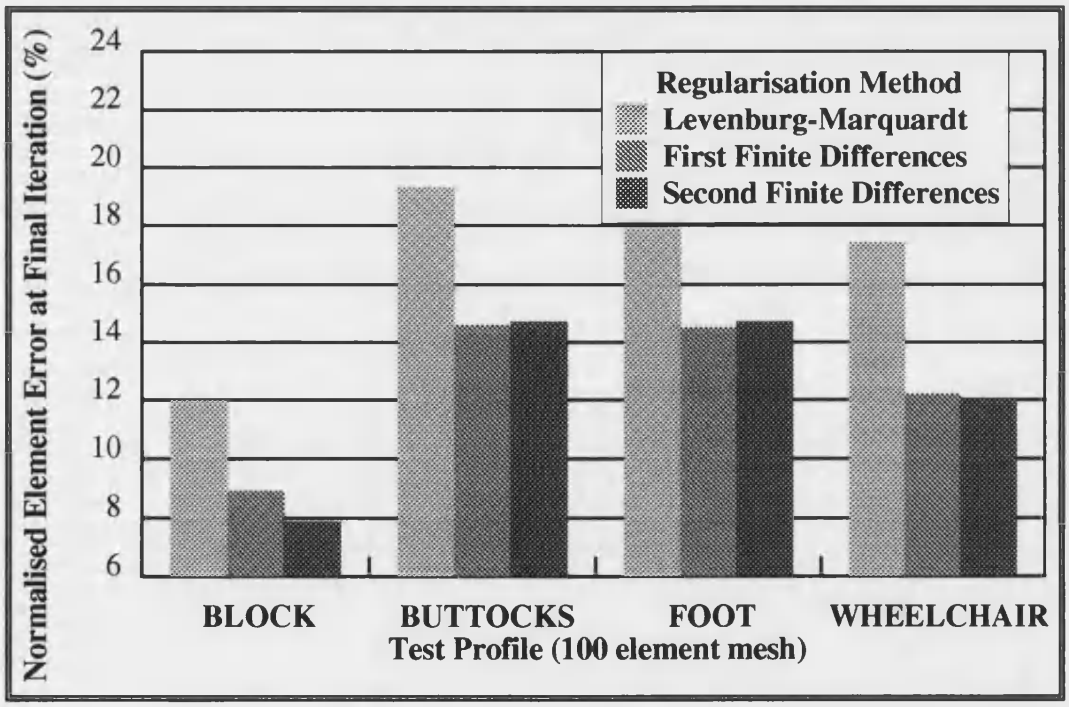


Figure 6.14 Regularisation method comparisons on the 100 element mesh with added voltage measurement error of 1%.

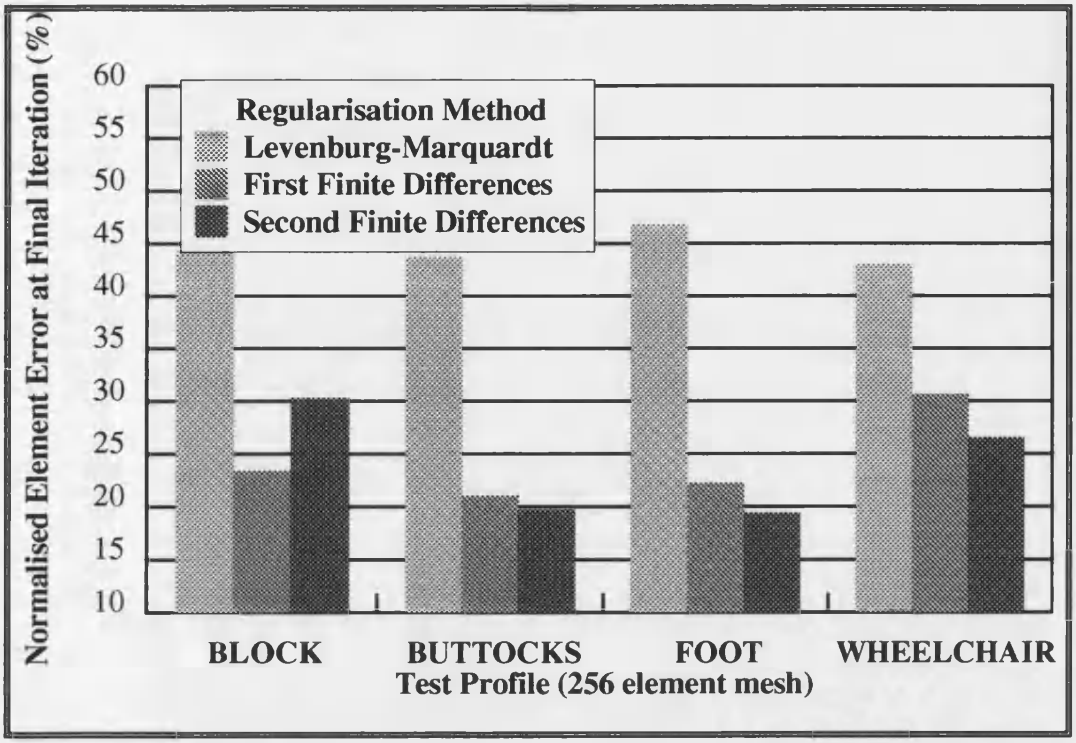


Figure 6.15 Regularisation method comparisons on the 256 element mesh with added voltage measurement error of 1%.

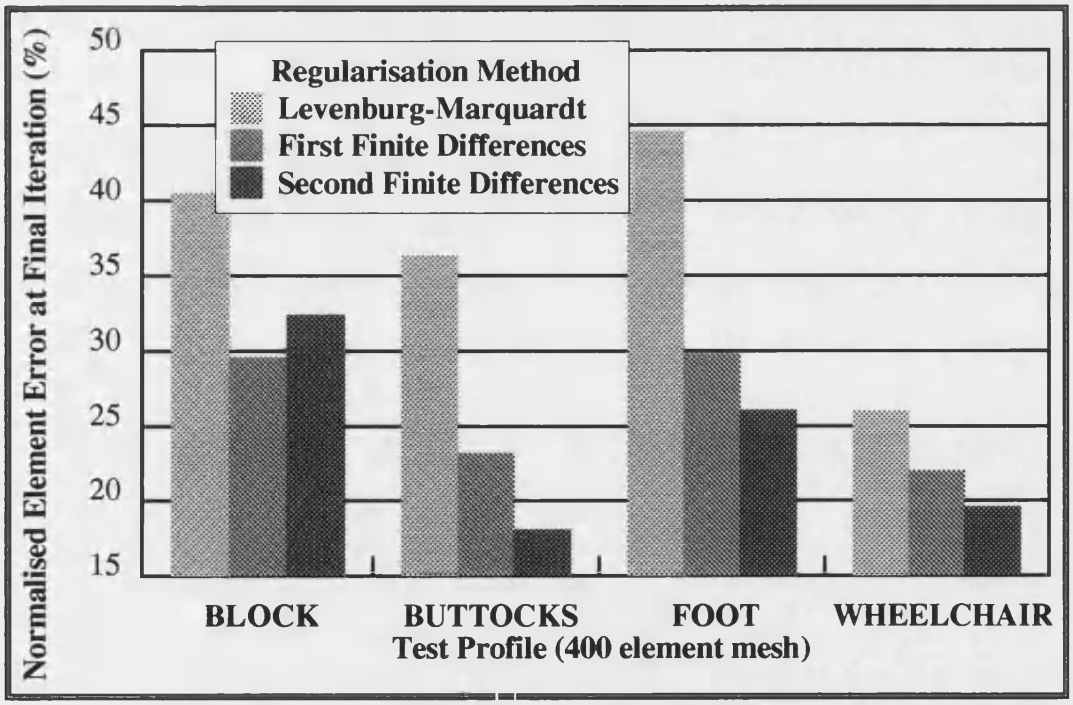


Figure 6.16 Regularisation method comparisons on the 400 element mesh with added voltage measurement error of 1%.

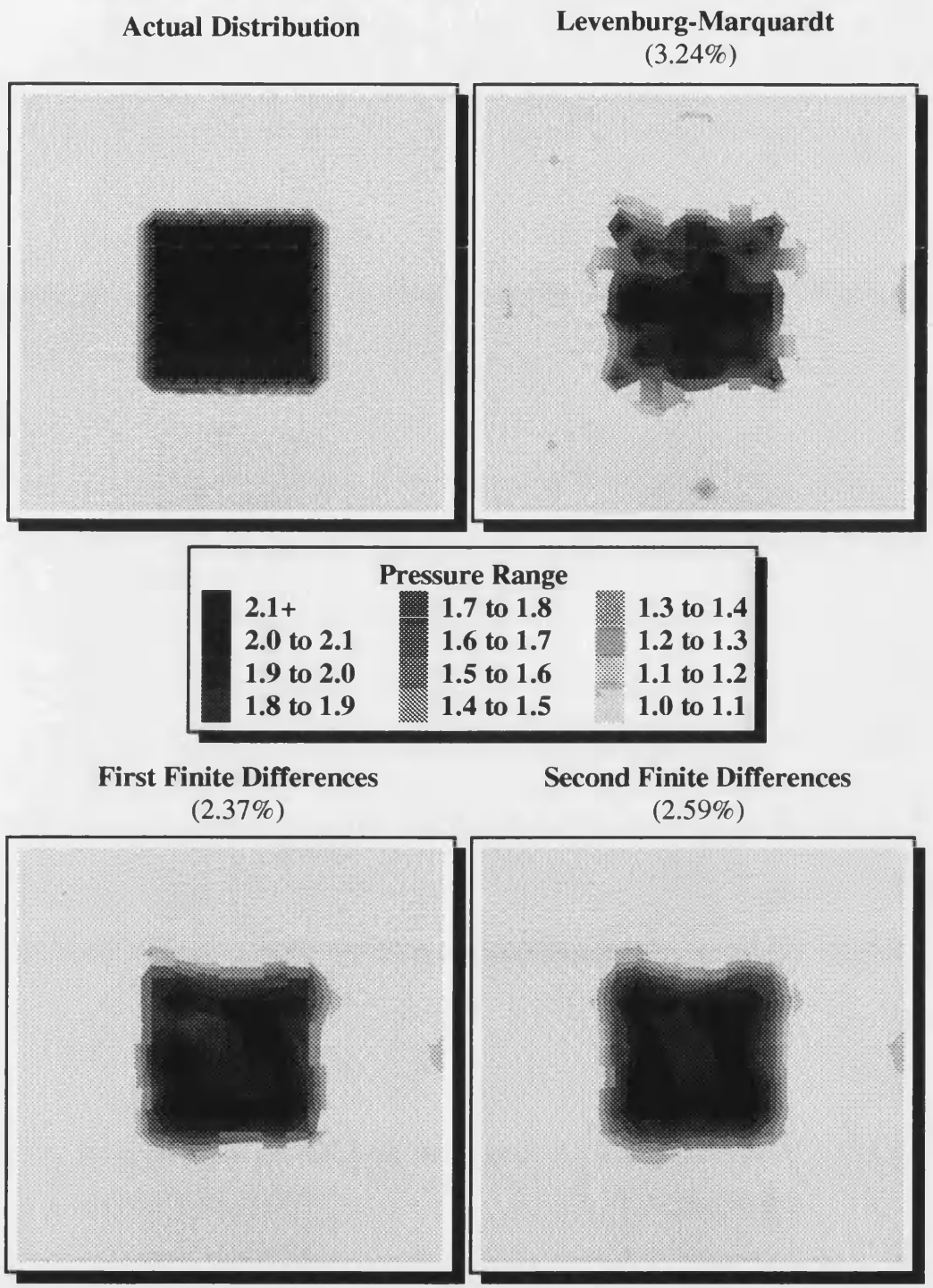


Figure 6.17 Test profile BLOCK on the 400 element mesh and the reconstructions obtained using the 3 different regularisation methods (1% voltage error). The final element errors are shown in brackets for each method.

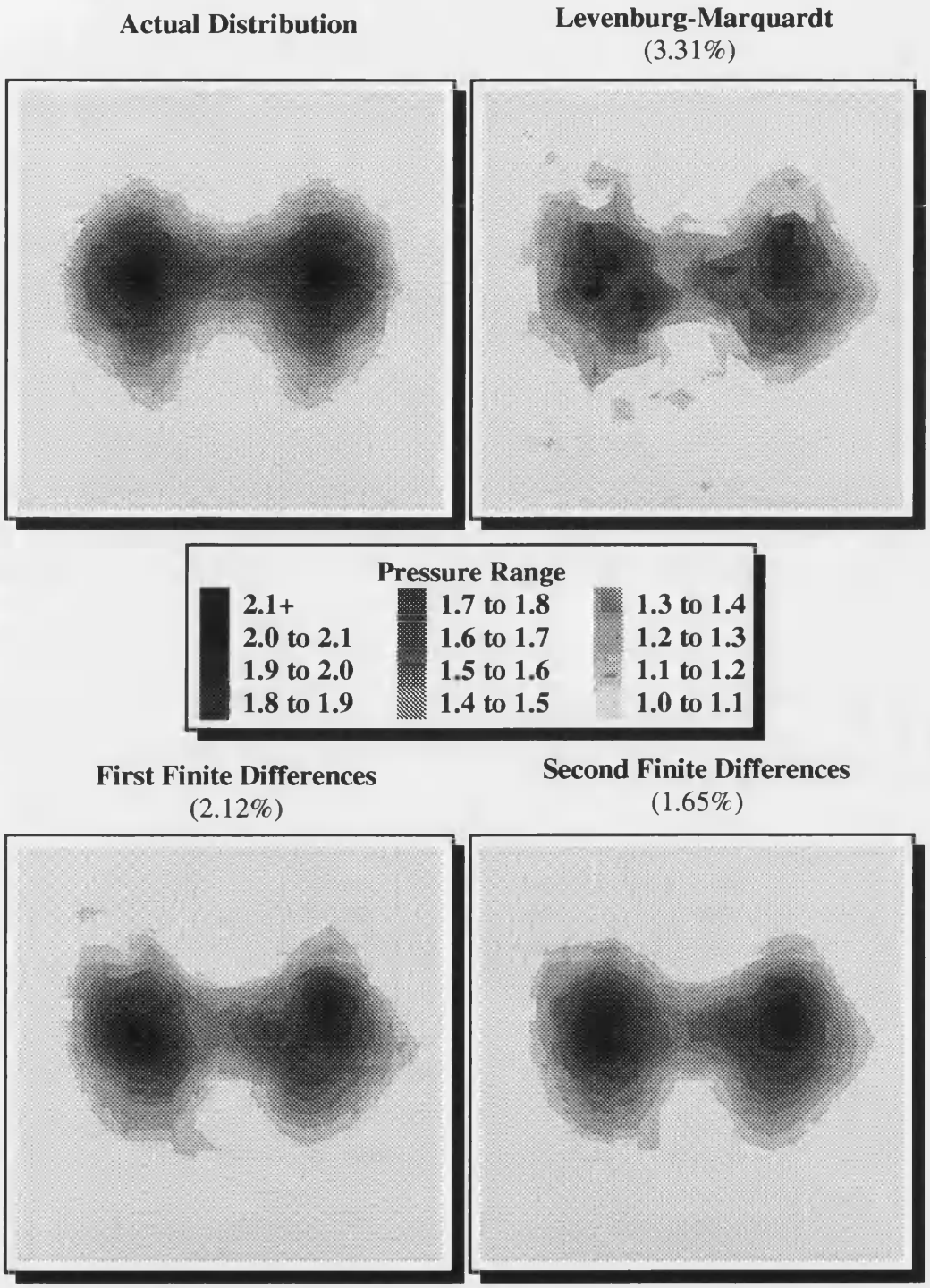


Figure 6.18 Test profile BUTTOCKS on the 400 element mesh and the reconstructions obtained using the 3 different regularisation methods (1% voltage error). The final element errors are shown in brackets for each method.

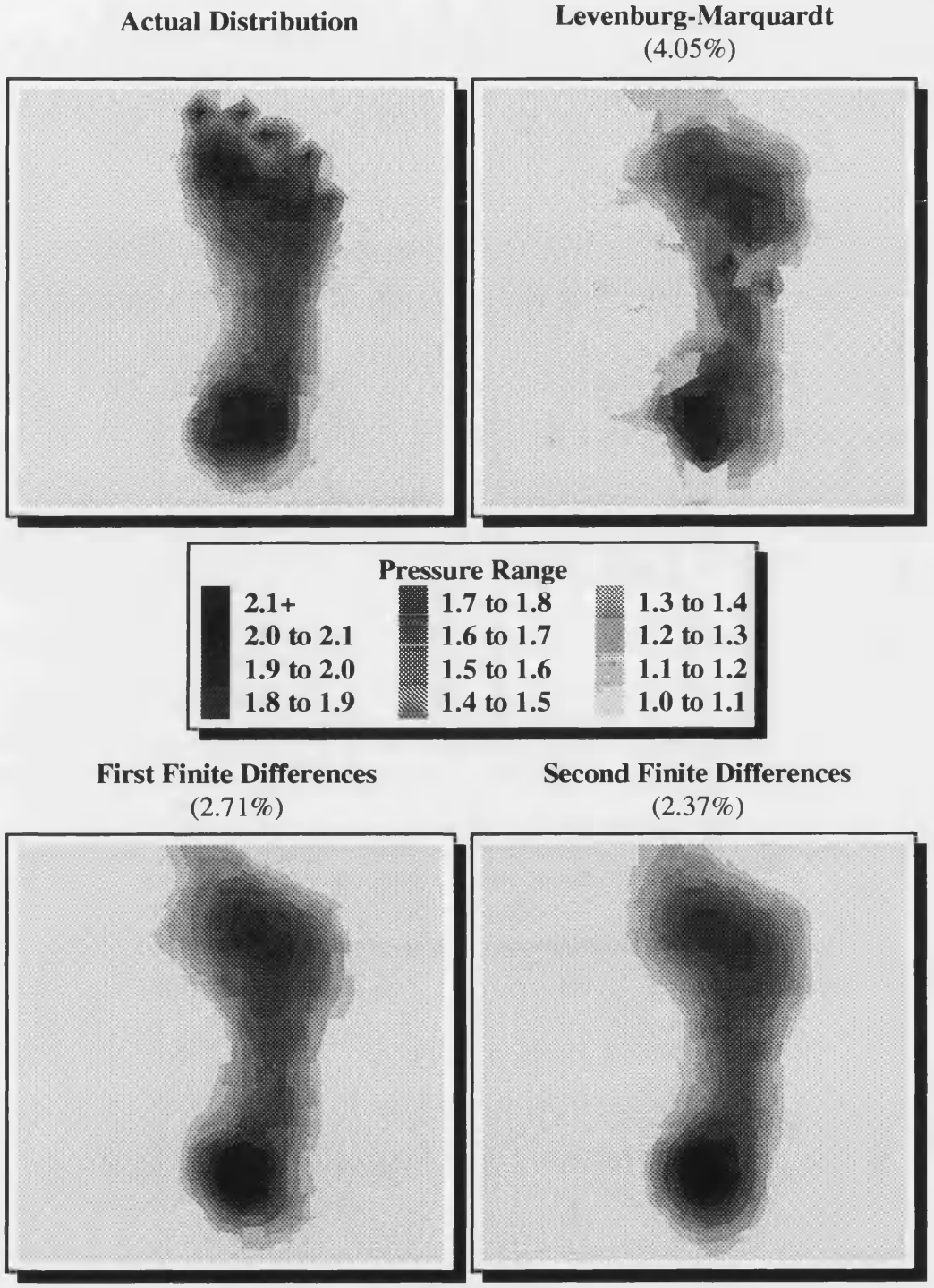


Figure 6.19 Test profile FOOT on the 400 element mesh and the reconstructions obtained using the 3 different regularisation methods (1% voltage error). The final element errors are shown in brackets for each method.

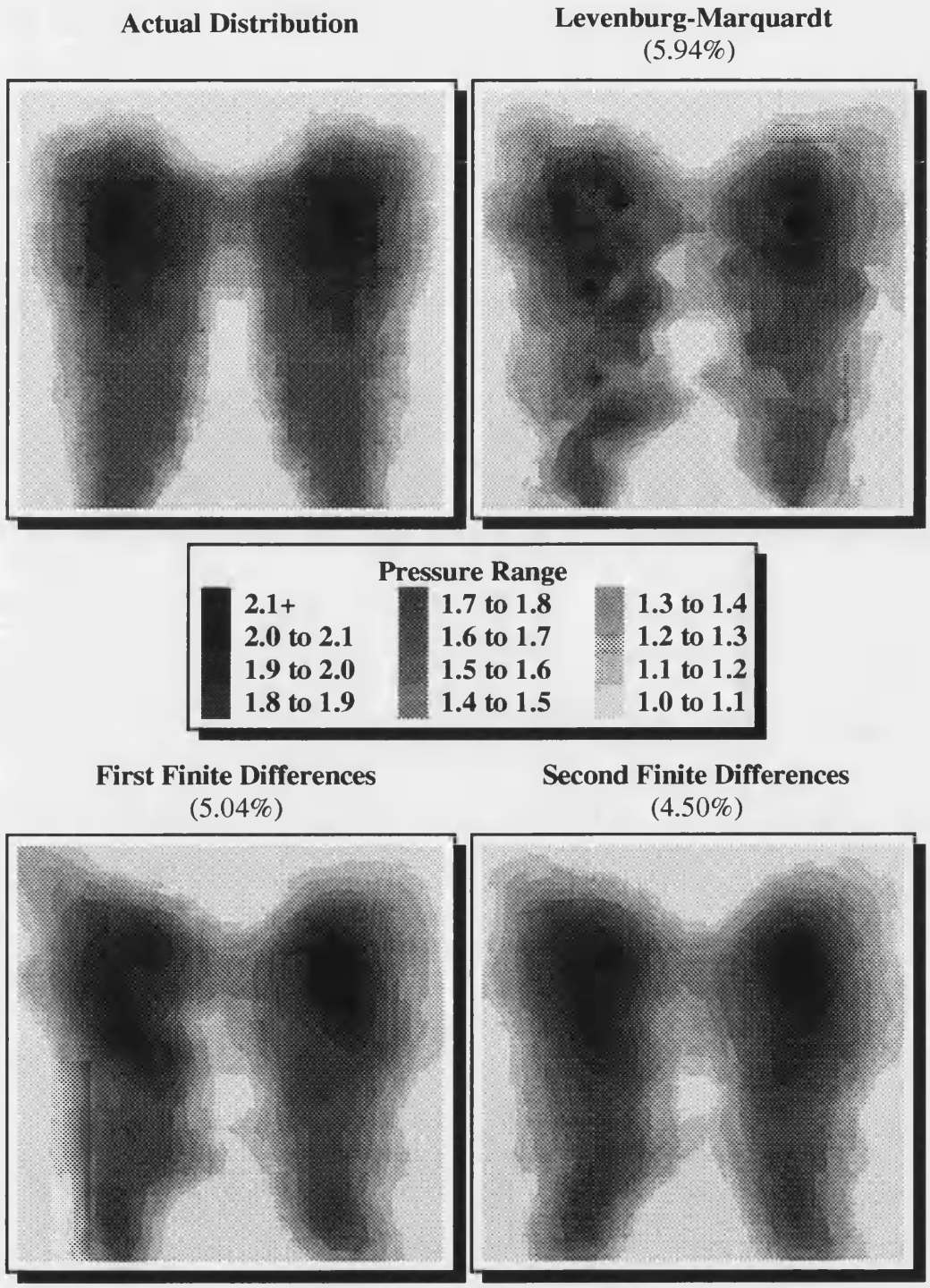


Figure 6.20 Test profile WHEELCHAIR on the 400 element mesh and the reconstructions obtained using the 3 different regularisation methods (1% voltage error). The final element errors are shown in brackets for each method.

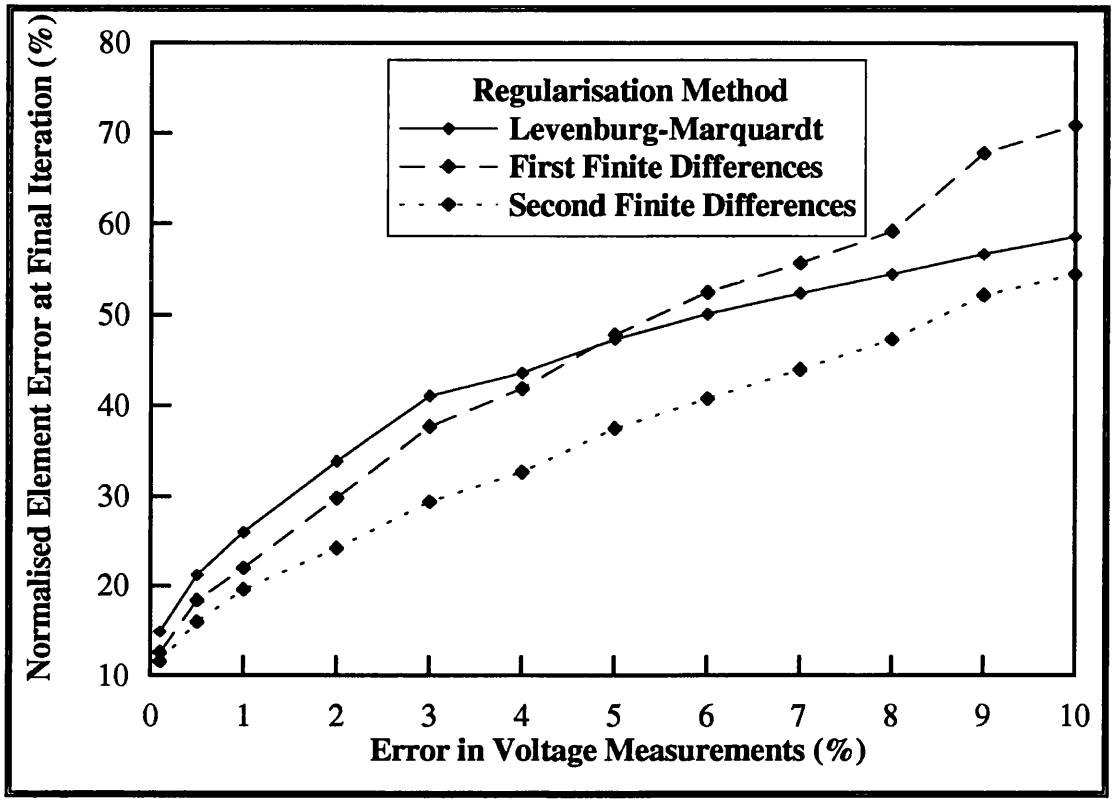


Figure 6.21 Reconstruction comparisons on test profile WHEELCHAIR, 400 element mesh, over a range of voltage measurement errors.

The first conclusion from these results is that the Levenberg-Marquardt regularisation always gives a reconstruction inferior to those produced by the other two. The second difference regularisation generally produced a lower element error than first difference regularisation. There are two notable exceptions, namely test profile BLOCK on the 256 and 400 element meshes. This is due to the greater smoothing of the second difference regularisation on the sharp edges that are present.

The regularisation method comparisons were checked over different noise levels. The error in the reconstruction over a range of voltage measurement noise levels is typified by test profile WHEELCHAIR on the 400 element mesh and is shown in Figure 6.21. The second difference regularisation is again the best over all noise levels. Even though the Levenberg-Marquardt regularisation gives a lower element error than the first finite difference regularisation at higher noise levels, the reconstructions it produced were not necessarily better. Figure 6.22 demonstrates this when the voltage error is 10%.

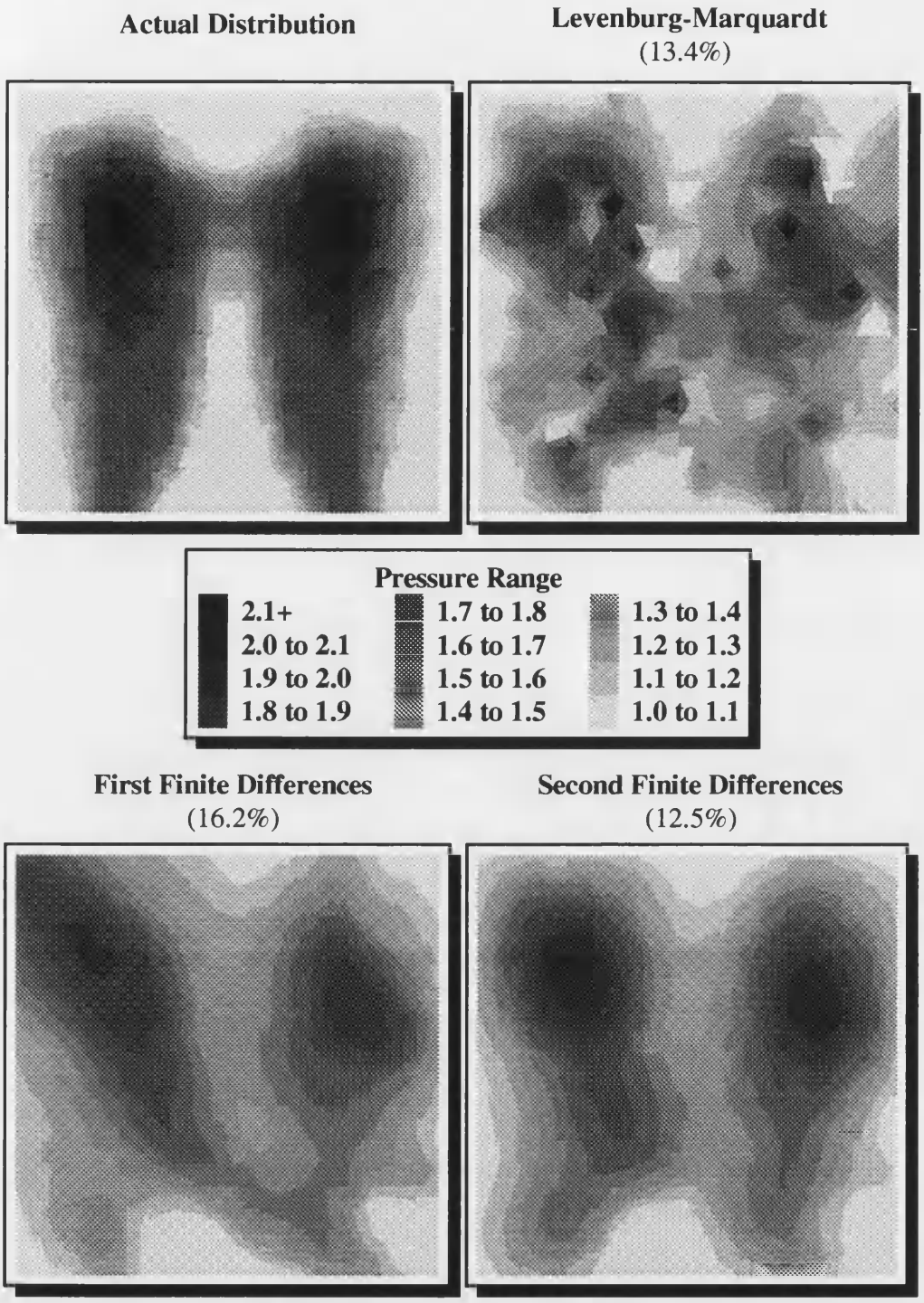


Figure 6.22 Test profile WHEELCHAIR on the 400 element mesh and the reconstructions obtained using the 3 different regularisation methods (10% voltage error). The final element errors are shown in brackets for each method.

The Levenberg-Marquardt method increases all the eigenvalues of $\mathbf{J}^T\mathbf{J}$ by λ, and so the conditioning of the matrix to be 'inverted' is improved. Similarly, the other two regularisation methods improve the conditioning of the equations being solved. Noise has less effect on the quality of the reconstruction as the degree of regularisation increases and so regularisation using second finite differences is the most effective. There is a stronger binding of one element to its neighbours when using penalty terms of higher orders, effectively smoothing the results. In essence, regularisation acts as a spatial low-pass filter and improves the results produced by the Gauss-Newton method, but at the expense of spatial resolution. This lower spatial resolution is apparent in Figure 6.19, where the toes cannot be differentiated from the rest of the foot.

6.3.5. Regularisation Parameter Variations

Five different methods to choose the regularisation parameter at each iteration were discussed in the previous chapter; these were the Marquardt, Davies, dwindling, constant and Moré regularisation parameter methods. In this section, the first three are compared against each other, followed by an examination of the latter two. Consideration is given to the speed of reconstruction and accuracy of the reconstruction.

Before comparing reconstructions by the Marquardt, Davies and dwindling parameter methods, the effect on the reconstruction of any predefined parameters needed by each method needs to be known. The Davies method is entirely automatic and does not require any predefined parameters. The Marquardt method requires a multiplier, v_m, a divisor, v_d and an initial regularisation parameter, λ^0, to be pre-set. The dwindling method only requires the latter two of these. Although there are a number of variables to be pre-set in these last two methods, reconstructions were fairly insensitive to variations in them. A conservatively large $\lambda^0 = 0.1$ was used on all test profiles and mesh sizes so that the first step is biased towards the steepest descent direction and far away from the Gauss-Newton direction. This is chosen as the steepest descent direction is superior to the Gauss-Newton direction on the first iteration, i.e. when far from the solution point. It was found that $\lambda^0 > 0.1$ gave no improvement nor change in the final reconstruction error. However, the number of iterations to the solution increased, albeit insignificantly. If this value was set too low, a degraded final reconstruction would occur, hence the conservatively large value. Variations in v_d and v_m gave small changes in the final reconstruction error. As these minor improvements are dependent on the pressure profiles and the mesh size used, it does not warrant optimisation due to the wide variety

of pressure distributions that the mat must reconstruct. The values suggested by Marquardt are used for all tests, namely $v_m = v_d = 10$.

Using the values mentioned, the Marquardt, dwindling and Davies methods were compared using the meshes and test profiles mentioned earlier in this chapter. The final reconstruction errors for the four test profiles on the 100 element mesh are shown in Figure 6.23 and for the 400 element mesh in Figure 6.24. The general trends are quite clear. The Davies method does not perform as well as either of the Marquardt or dwindling parameter methods. The Marquardt method gives the lowest reconstruction error and is slightly better than the dwindling method. However, the Marquardt method requires the greater number of iterations of these two. This can be seen in Figures 6.25 and 6.26.

For a real-time system, the extra accuracy may be traded off for a faster reconstruction. In this case the dwindling parameter method would be a preferable choice. Alternatively, the convergence tolerance could be increased for the Marquardt method. A typical reconstruction history is shown in Figure 6.27 for the BUTTOCKS test profile on the 100 element mesh. This figure clearly shows that the dwindling method is essentially a Marquardt method which has been terminated early (at the point at which divergence would first have occurred). The figure also shows that by lowering the convergence tolerance level, the Marquardt iterative process could be terminated earlier. This would require correctly estimating the tolerance level at an appropriate level so that the process is terminated early. This varies from one pressure distribution to the next, but would be a good compromise between accuracy and time to reconstruction. The dwindling method is more reliable for fast reconstruction as divergence usually occurs after 5 or 6 iterations.

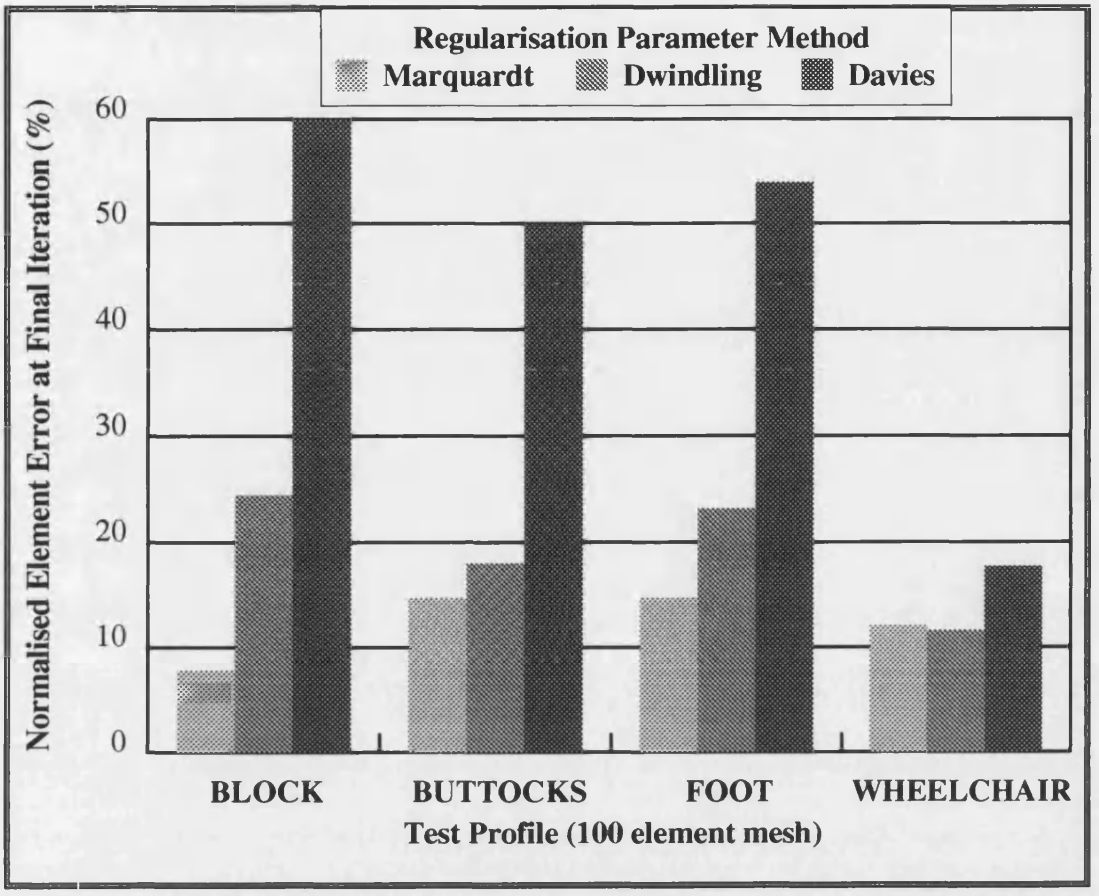


Figure 6.23 Reconstruction errors on 4 different 100 element mesh test profiles using regularisation with second difference penalty terms. The added voltage measurement error is set at 1%. Comparisons are made between three different methods for choosing the regularisation parameter.

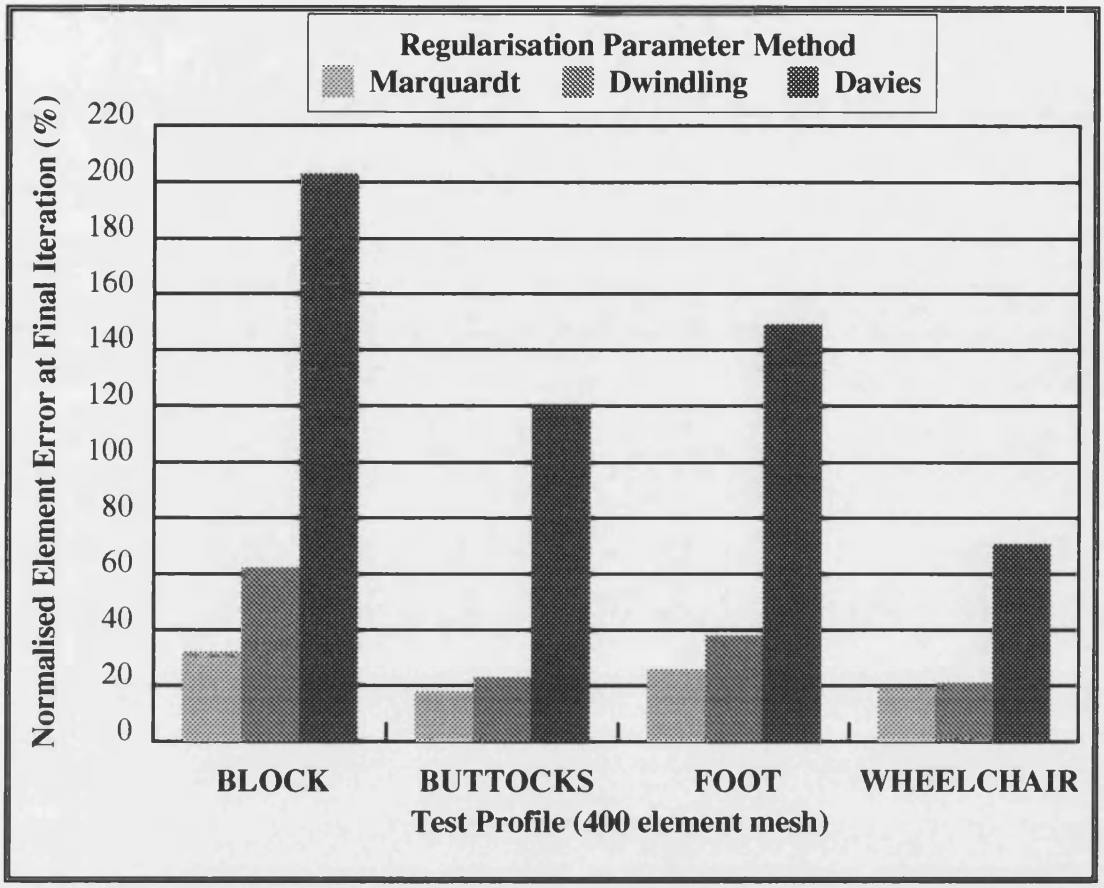


Figure 6.24 Reconstruction errors using the same set-up as Figure 6.23 but on the 400 element mesh.

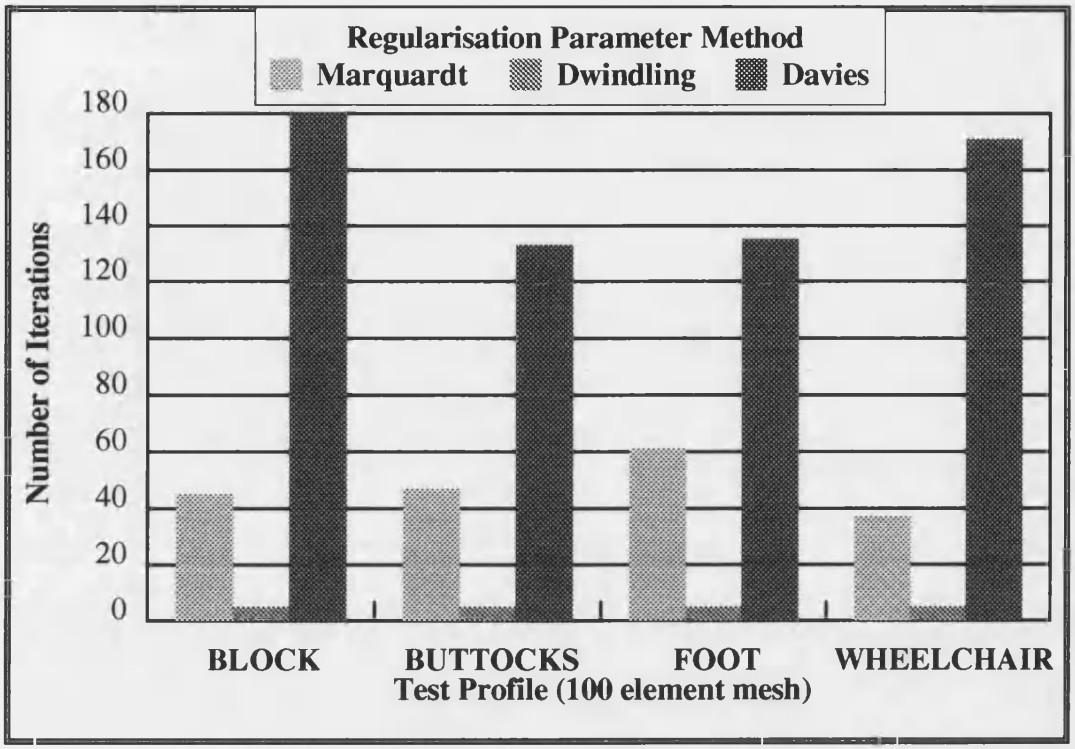


Figure 6.25 Number of iterations to convergence for the 100 element test profiles of Figure 6.23

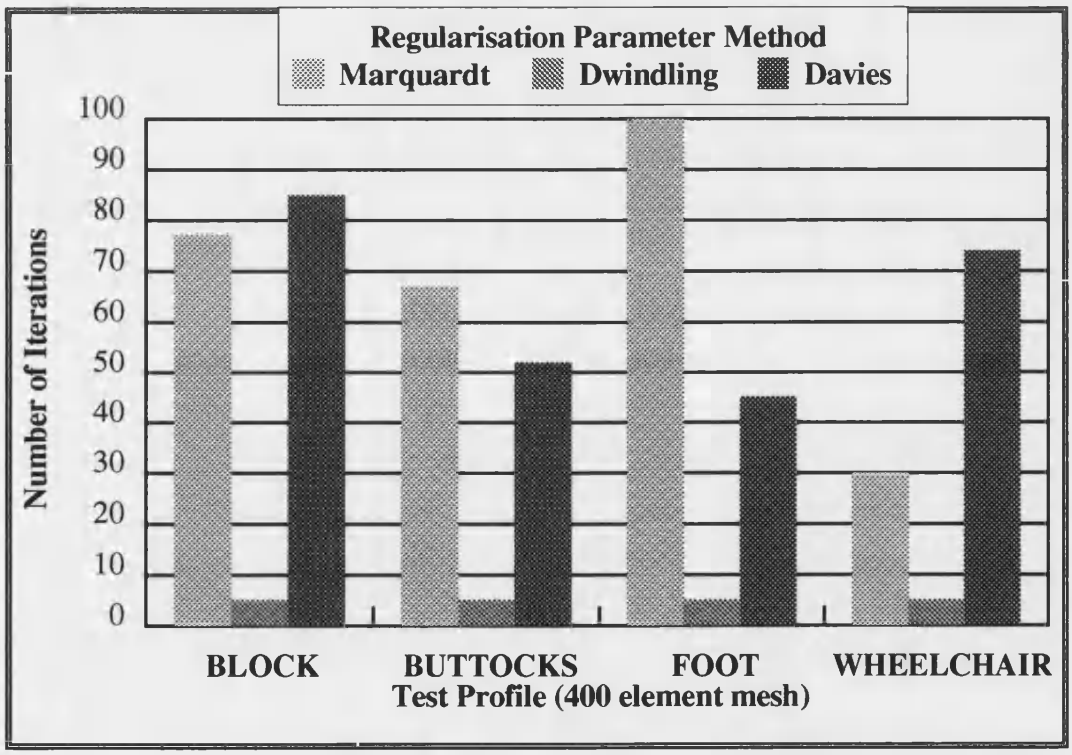


Figure 6.26 Number of iterations to convergence for the 400 element test profiles of Figure 6.24

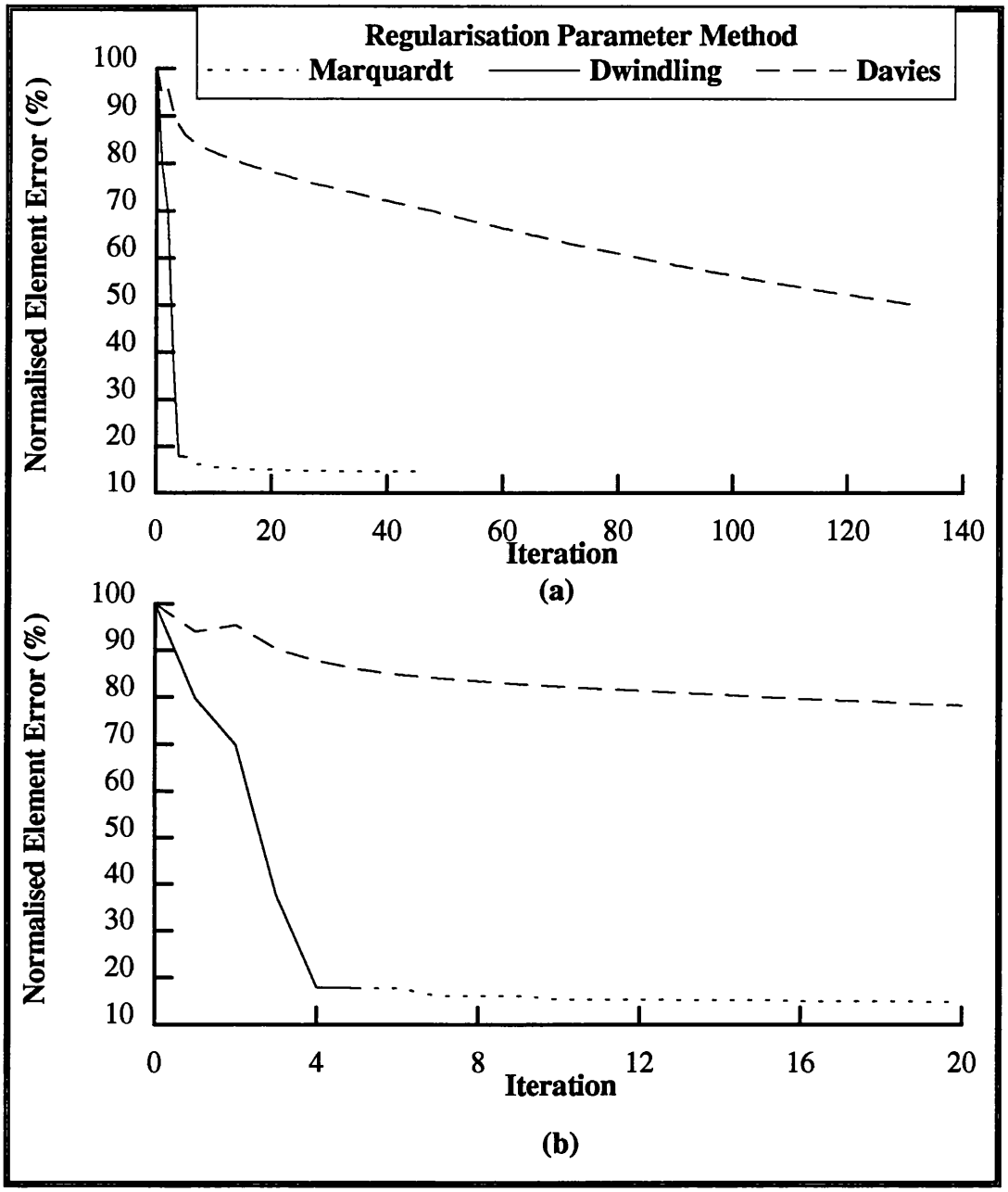


Figure 6.27(a) Reconstruction progress of the BUTTOCKS test profile used in Figure 6.23.
(b) The first 20 iterations of (a).

6.3.5.1. Using the Constant Regularisation Parameter

Using a constant value is a simple approach for choosing the regularisation parameter. This also proved to be an effective approach, but with a limitation. In contrast to the

previous three dynamic methods, the reconstruction is very sensitive to the choice of the regularisation parameter. The 4 test profiles on the 100 element mesh were reconstructed using the constant regularisation parameter method. The final normalised reconstruction errors are shown in Figure 6.28 for a variety of regularisation parameters. This graph highlights the problem with the constant regularisation parameter method. The optimum regularisation parameter value is dependent on the pressure distribution being reconstructed. With two of the test distributions, BLOCK and WHEELCHAIR, the optimum constant regularisation parameter gives a reconstruction that is slightly more accurate than the Marquardt method of Figure 6.23. However, the constant regularisation parameter method is quite sensitive to the choice of λ. The Marquardt method is the preferred choice as it is more consistent at providing low errors in the reconstruction.

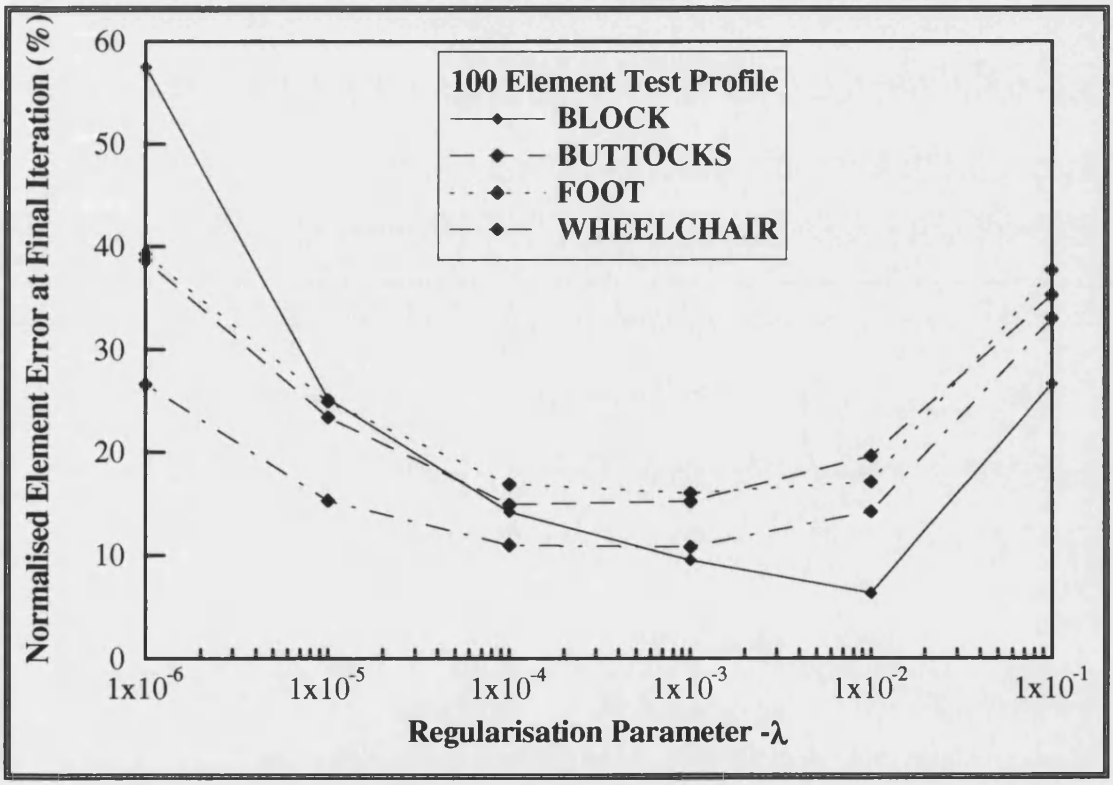


Figure 6.28 Reconstruction errors with various constant regularisation parameters. Regularisation parameter method excluded, the set-up is the same as Figure 6.23.

6.3.5.2. Using the Moré Regularisation Parameter

Moré[76]'s trust region approach to finding λ is implemented with the Levenberg-Marquardt method in a library routine called LMDER in MINPACK[85]. The source code is obtainable from a public domain software base on the internet (*http://www.netlib.org/minpack*). Unfortunately this approach is not well suited to the problem in hand. For the extensive test distributions tried, the reconstructed results rarely improved on the Levenberg-Marquardt algorithm with λ chosen by the other algorithms. The size of the trust region is dictated by the variable r_t and an initial estimate for it is required by the algorithm. The results produced were very sensitive to the initial trust region size. For recognisable reconstruction images to occur, especially with the larger grids, the initial trust region had to be much smaller than the recommended range. A small trust region is equivalent to a large regularisation parameter. With larger trust regions, the search direction tended towards the Gauss-Newton direction. The net effect was convergence to false minima as not enough regularisation was introduced. The regularisation parameter is based on an estimate of the trust region size obtained from the previous iteration. Even when a good initial trust region size was used, the algorithm tended to over-expand the trust region for the following iteration. This constant bias towards the Gauss-Newton direction, usually led to reconstructions converging towards false minima. The conditions under which this algorithm proved to be superior was when the simulated voltage measurement error was too low for practical purposes. The conclusion must then be that the trust region approach is best for small residual problems.

6.3.5.3. Summary

A general trend can be observed amongst the different methods for selecting λ. When any of them used an initial λ that was very small, the procedure would not reconstruct any pressure distribution particularly well. The necessary larger values of λ^0 move the search direction away from the Gauss-Newton direction, indicating the importance of regularising the Gauss-Newton method for this ill-conditioned problem. The most consistently accurate method proved to be the Marquardt method. Although, the trust region method is a more modern approach, and is very well regarded in the optimisation literature, it failed to produce useful results with any degree of practical voltage measurement error. Moré's trust region approach was too heavily biased towards the Gauss-Newton direction for successful reconstruction in the presence of noise.

Some systems require a fast reconstruction, e.g. a real-time pressure scanner. In this case, a minimal number of iterations to a reasonably accurate solution is required. The early termination of the Marquardt method, that is, the dwindling parameter method, provides a good balance between speed and accuracy in the solution. The method usually stops after 5 or 6 iterations. This is with the initial estimate set at the 'no-load' pressure distribution. In a dynamic real-time pressure scanner, the previous pressure distribution can be used as the initial estimate for the following scan. In this situation, slightly fewer iterations will be required to arrive at a reasonably accurate solution.

The table in Figure 6.29 summarises the times taken to solve the regularised Gauss-Newton test problems using the dwindling regularisation parameter on a Sun 20/51.

Mesh Size	Time (s)	Iterations
64	0.6	5
100	1.9	5
256	20.2	6
400	63.5	6

Figure 6.29 The times and number of iterations required for reconstruction of the different sized test profiles on a Sun 20/51. Reconstruction is with the second finite difference regularisation and dwindling regularisation parameter method.

6.3.6. Scaling

The results shown in this thesis use well-scaled equations. That is, the resistivities are all of similar magnitude and are in the region around unity. Although, a pressure sensitive material has not been manufactured, the no-load resistivity can be manufactured to any value by varying the degree of carbonisation of the fibres. No scaling is used in this thesis as scaling the well-scaled equations actually gave worse results. However, a few experiments with resistivity values in the region of 1000 required scaling for successful reconstruction. Woo[83] implemented the Marquardt scaling and also found it very important to improve the solution accuracy. When a pressure sensitive material with known resistivity characteristics is manufactured, scaling may well have to be examined more closely.

6.3.7. Effects of the Bounds Constraints

Reconstruction errors with and without the bounds constraints are shown in Figure 6.30. The bounds constraints have a noticeable improvement on the success of the reconstruction. These results on the 400 element mesh using second difference regularisation are typical of other mesh sizes and reconstruction algorithms. As such all reconstruction algorithms results shown in this thesis have used these simple bounds constraints.

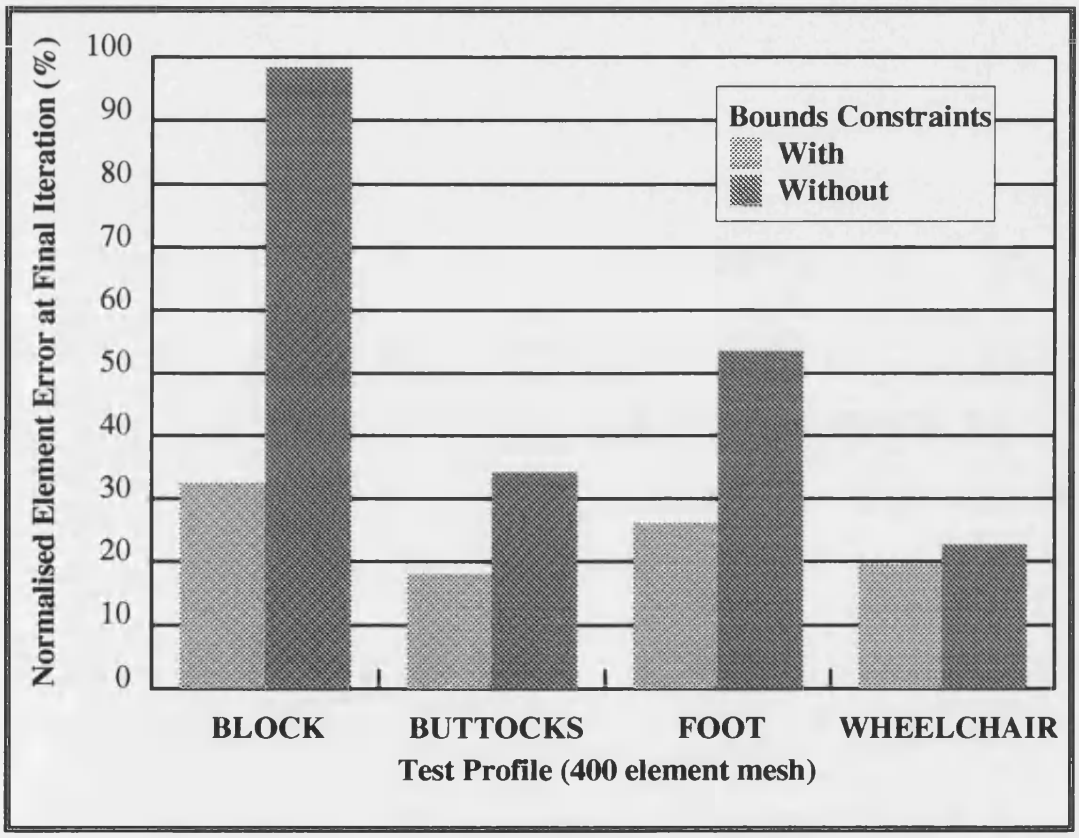


Figure 6.30 Reconstruction errors with and without bounds constraints on the 400 element test profiles using second difference regularisation. Voltage measurement error is set at 1%.

6.4. Conclusions

The Gauss-Newton and derived methods were examined for suitability of solving the pressure EIT problem. The damped and undamped Gauss-Newton methods work very well in ideal conditions of zero measurement error, but the Gauss-Newton direction

tends to false minima with noise contaminated data. The problem is completely altered from a zero residual problem to a large residual problem when realistic measurement errors are introduced. Regularisation is needed to tackle the solution when the residuals are no longer zero. It was found that regularisation based on second difference penalties was best in these circumstances. The regularisation parameter that consistently gave good results was the Marquardt choice. For a compromise between accuracy and speed of solution, the dwindling parameter method should be used. A number of algorithmic and software implementational techniques have also decreased the time of each iteration. The use of a single earthed node removed the need for recalculation of the master stiffness matrix inverse for each and every projection angle.

The voltage measurement errors usually associated with medical EIT applications will not be as large when applied to a resistive mat; electrode placement will be permanently fixed and currents will stay two-dimensional. The problem is ill-conditioned, which causes the solution to be highly sensitive to any voltage measurement errors. However, the bounds constrained regularised Gauss-Newton method reconstructs the pressure distributions very well even in the presence of realistic noise levels.

7. Electrode Positioning

This chapter demonstrates the reconstruction improvements that can be obtained with the use of electrodes distributed throughout the pressure sensitive mat, as opposed to attaching the electrodes entirely on the mat's edge. The optimal positioning of the electrodes is studied.

The reconstruction accuracy is dependent on a number of factors. These include the reconstruction algorithm employed, the current patterns applied to the object, the pressure distribution itself, the positioning of the electrodes and voltage measurement and other system errors. This chapter examines the effects of the electrode positioning. For an objective comparison between a variety of electrode positions, the aforementioned variables will be kept constant. However, the applied current patterns depend on the electrode positioning and cannot be kept the same for two different sets of electrode positions. To obtain the most objective comparisons, the so-called optimal current patterns are used. These current patterns are optimal in that they are the best currents to apply for minimum degradation in the reconstruction in the presence of noise. The first part of this chapter examines optimal current patterns. The second part of this chapter studies various electrode positions using the optimal current patterns. The study analyses distinguishability and the Jacobian matrix conditioning, showing the effects on reconstruction error.

7.1. Optimal Current Patterns

7.1.1. Projections

A single projection or current pattern is simply the set of electrode currents applied to the mat. In EIT, a set of voltage measurements is then taken from the electrodes for each projection. A number of different current patterns are applied to the mat in order to gain enough information for the reconstruction to succeed. Many types of current patterns can be applied to the mat. The neighbouring current patterns of Brown[86] are commonly used. Current is injected into one electrode and extracted from a neighbouring electrode. This is repeated for all adjacent pair electrodes. Hua[87] presents and compares other current patterns which involve two current carrying electrodes per projection. However, it is the current patterns which utilise all electrodes in each projection that produce the best results.

7.1.2. Distinguishability

The optimal current patterns involve applying currents to all available electrodes and were proposed by Isaacson[88]. Considering two different resistivity distributions, ρ_1 and ρ_2, the electrode voltages resulting from an applied current pattern, **c**, are $\mathbf{v}(\rho_1)$ and $\mathbf{v}(\rho_2)$ respectively. The two distributions are said to be distinguishable when

$$\delta = \frac{\|\mathbf{v}(\rho_1) - \mathbf{v}(\rho_2)\|}{\|\mathbf{c}\|} \tag{7.1}$$

is greater than the system error measurement tolerance or noise, ε. The distinguishability, δ, is a function of the applied current pattern as well as the two resistivity distributions and is merely a measure of the voltage differences at the electrodes normalised by the applied currents. Optimal current patterns maximise the distinguishability so that $\delta > \varepsilon$ even when the two resistivity distributions are very close. This is important for the least squares reconstruction method. If $\mathbf{v}(\rho_1)$ is the vector of measured voltages and the estimate at any given iteration is $\mathbf{v}(\rho_2)$, then the numerator in equation (7.1) is the basis of the reconstruction method, see equation (5.1). As soon as $\varepsilon > \delta$, then voltage measurement error will dominate the next iteration in the reconstruction update. By maximising δ, the signal to noise ratio is increased and the effect of measurement error is reduced. Optimal currents are thus optimal in that they minimise the error introduced into the reconstruction method simply by maximising the distinguishability.

7.1.3. Calculation of Optimal Currents

The calculation and implementation of the optimal currents is a CPU intensive task and is described in this section. A few reconstructions comparing results using optimal current and other current patterns follows.

7.1.3.1. Difference Matrix Calculation

The equation $\mathbf{v}(\rho) = \mathbf{R}(\rho)\mathbf{c}$ describes Ohm's law, where **R** is the transfer impedance matrix for some resistivity distribution ρ. If all nodes in the FEM mesh were used as electrodes, **R** would then be the generalised inverse of the admittance matrix, **Y**. For the same current pattern applied to two different distributions, the distinguishability equation (7.1) can then be rewritten as

$$\rho = \frac{\|\mathbf{Dc}\|}{\|\mathbf{c}\|} \tag{7.2}$$

where **D** is the difference matrix,

$$\mathbf{D} = \mathbf{R}(\rho_1) - \mathbf{R}(\rho_2) \tag{7.3}$$

Isaacson[88] showed that the maximum distinguishability is obtained when the currents are the eigenvectors of **D** relating to its largest eigenvalues. The entire set of eigenvectors then make up the optimal current patterns. Note that the optimal currents are a function of the unknown resistivity distribution and as such different optimal currents exist for different resistivity distributions.

7.1.3.2. Finding the Eigensystem

The matrix, **D**, is a square symmetric matrix and so the eigensystem can be found from the Singular Value Decomposition (SVD). The SVD decomposes the $e \times e$ **D** matrix as follows[89]

$$\mathbf{D} = \mathbf{U}\Sigma\mathbf{Z}^{\mathrm{T}} \tag{7.4}$$

where Σ is a diagonal matrix of decreasing positive singular values, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_e$ and **U** & **Z** are orthogonal matrices. The columns of **U** and **Z** are respectively the left and right singular vectors. As **D** is square and symmetric the singular values are the same as the eigenvalues and the left and right singular vectors are the same.

Note that as **R**, and therefore **D**, is singular of rank e-1, where e is the number of electrodes, the last singular value is zero. The optimal currents are the first e-1 left singular vectors. The following section explain how the **D** matrix is obtained.

7.1.3.3. Transfer Impedance Matrix Calculation

The transfer impedance matrices of two resistivity distributions must be found to calculate **D**. The admittance matrix is singular and so the transfer impedance matrix cannot be obtained by simply inverting it. Gisser[90] describes an experimental method to find **D**. This involved an iterative process of repeated measurements from the resistivity domain. Hua[91] overcame the time consuming nature and practical difficulties of this

process by suggesting a direct method to find **R**. This faster, as well as more accurate method of finding **D**, can be summarised as follows.

- Choose a set of independent current patterns or bases.
- Inject these current patterns into the actual resistivity distribution, ρ_1 and measure the voltage response.
- Find the transfer impedance matrix, $\mathbf{R}(\rho_1)$ using Ohm's law, Kirchoff's current law and the fact that the matrix is indefinite. An example is given in Appendix A3 for the so-called diagonal current patterns.
- Repeat using the same current patterns for the resistivity estimate (usually the previous iteration in the Gauss-Newton based methods) to find $\mathbf{R}(\rho_2)$. This time, however, the voltage response is obtained from the FEM.
- Matrix **D** is the difference between $\mathbf{R}(\rho_1)$ and $\mathbf{R}(\rho_2)$.

7.1.3.4. Optimal Currents and Synthesised Voltages

Injecting optimal currents into the domain necessitates a programmable current source on each and every electrode. This can be avoided by using a current basis that uses a single current source, e.g. the diagonal current patterns detailed in Appendix A3. Optimal currents are calculated from **D**, and because **R** is known, the voltage response to the optimal currents can be synthesised because $\mathbf{v} = \mathbf{Rc}$ [91]. The optimal currents' better distinguishability give superior reconstructions. So, even though diagonal current patterns are injected into the domain, the optimal currents and corresponding synthesised voltages can be used for the reconstruction process.

True optimal currents involve calculating and injecting optimal currents into the mat at the start of each and every iteration of the Gauss-Newton method. Each calculation requires the time consuming SVD. The reconstruction method is a fairly lengthy procedure in itself. Thus any single pressure distribution reconstruction requires the load to remain static for many seconds. This is likewise in the more common EIT applications and no known practical systems employ the full use of optimal currents. Some of the leading EIT researchers use spatial trigonometric current patterns[50][90]. Trigonometric currents turn out to be the optimal currents when considering a single, centrally located conductivity disturbance in a circular domain using equally spaced electrodes around the domain's edge. These patterns have proved to be very good in their applications and do not necessitate time consuming calculations in order to obtain the more ideal optimal currents. This approach requires the same single current pattern for all iterations.

However, as the electrodes are not evenly distributed around the edge of the pressure sensitive mat, spatial trigonometric currents cannot be applied in this case. Instead, a similar approach, detailed in the next paragraph, is used.

Diagonal currents are injected into the mat and voltage measurements taken for the test resistivity distribution. The optimal currents are calculated using these voltages and those from the initial estimate used in starting the regularised Gauss-Newton method. A sensible initial estimate is the constant resistivity distribution arising from the no-load pressure. This optimal current pattern and corresponding synthesised voltage measurements are used throughout the reconstruction. The current is then only optimal for the first iteration and near optimal for the remaining iterations. This has the advantage of a minimal time penalty incurred for each reconstruction.

7.1.4. Reconstruction Results using Optimal Currents

EIT researchers have always applied optimal currents to electrodes attached to the edge of a domain. Results showing the optimal currents' superiority over other current patterns using these edge electrode configurations already exist, for example Hua[49]. The reconstruction results in this section show optimal current improvements over other current patterns for an internal electrode configuration.

The optimal currents using diagonal currents, as described above, have been implemented in Fortran. The SVD is calculated using the NAG[55] library routine *f02wef*. All results are obtained from the computer simulation of the mat using the FEM. These optimal currents are injected into the mat and the resulting electrode voltage measurements are used for all iterations. The non-optimal current patterns with which comparisons are made are the diagonal currents. 32 electrodes are used and so the maximum of 31 independent projections for both current patterns are then fairly compared. The current patterns in Section 5 use fewer (16) projections which results in less maximally independent information, leading to slightly inferior reconstructions.

Many experiments on different sized meshes, using both optimal and diagonal current patterns, showed that the optimal currents always produced superior results. A selection of reconstructions for the 400 element mesh are shown in Figures 7.1 and 7.2. The test distributions and electrode positioning are those portrayed in Section 6.3.1. These typical results show a lower element error at reconstruction and visible superiority of the

optimal current patterns. One observation worth noting is the sharpening of the edges in the least smooth test, BLOCK.

The dwindling rather than the superior Marquardt regularisation parameter method was used. This is done for a secondary purpose; to show the quality in the reconstruction using this prematurely terminated, but faster approach. Despite errors in the reconstruction, the results are good and very usable. As expected, the optimal currents produced a higher distinguishability than those from the diagonal currents. In conclusion, optimal current theory can be applied to internal electrode configurations to produce improved reconstructions.

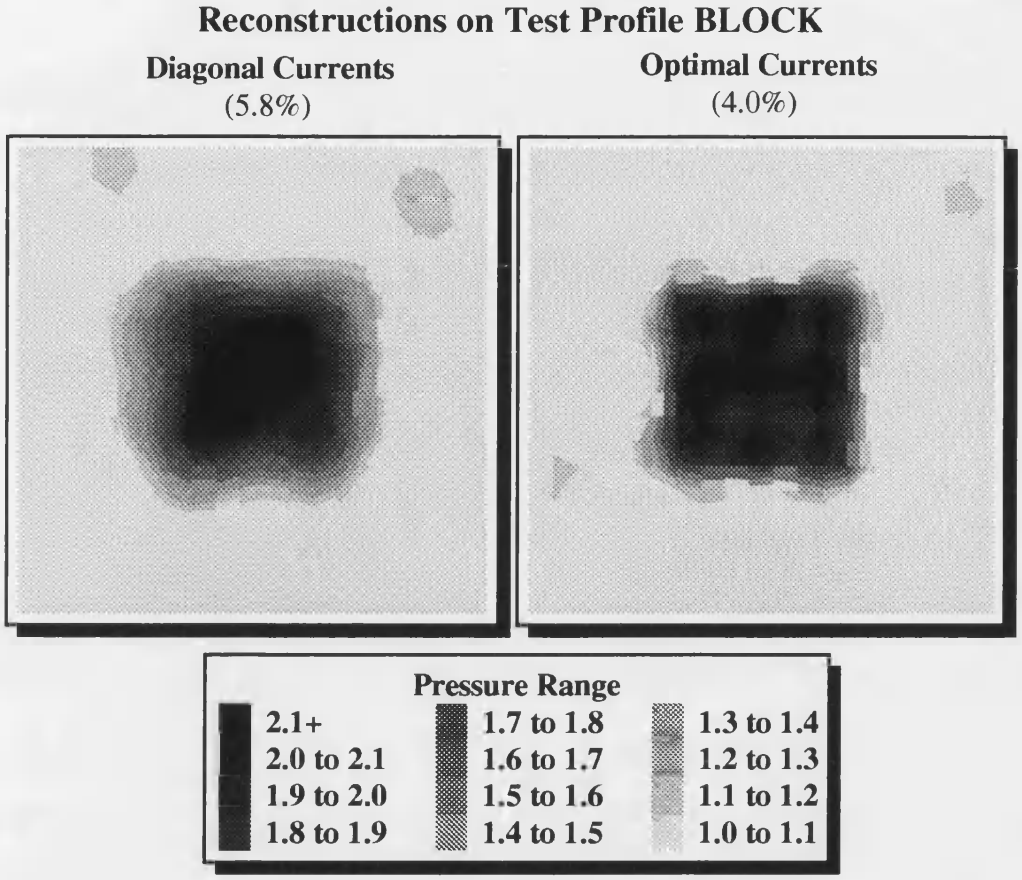


Figure 7.1 Diagonal and optimal current reconstructions comparisons using test profiles BLOCK and BUTTOCKS on the 400 element mesh. Added voltage error is 1% and the final element errors are shown in brackets.

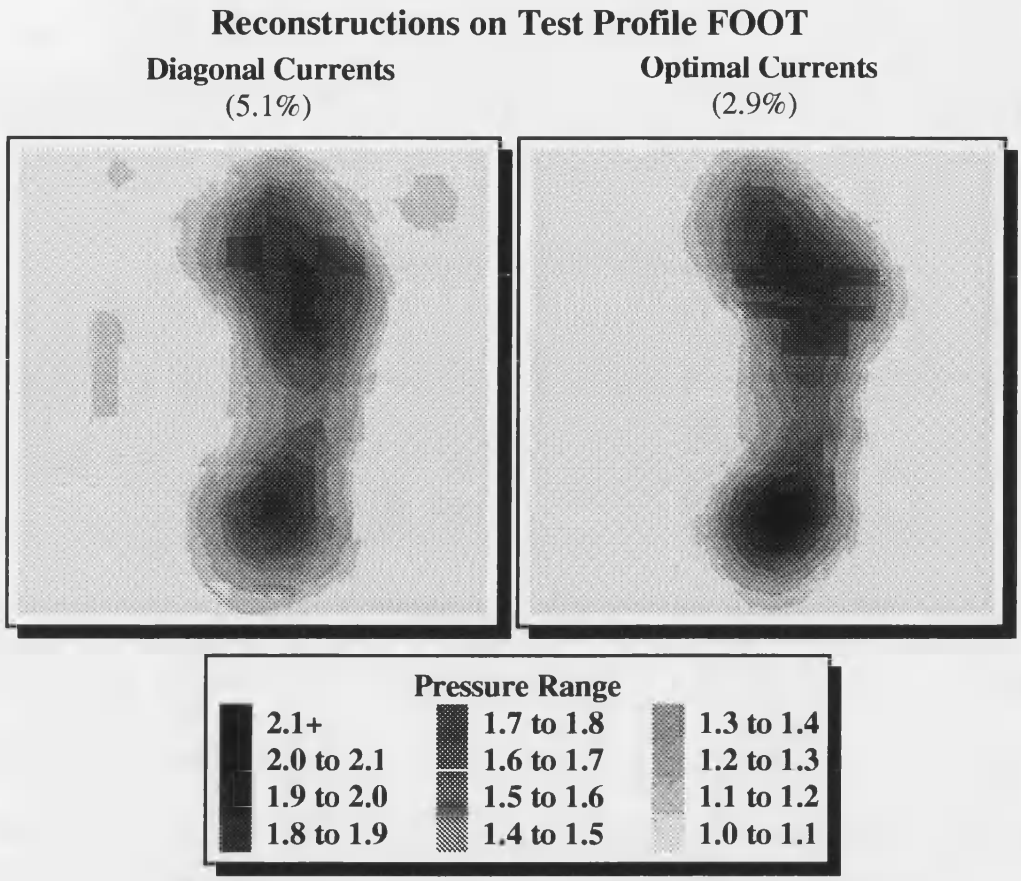


Figure 7.2 Diagonal and optimal current reconstructions comparisons using test profiles FOOT and WHEELCHAIR on the 400 element mesh. Added voltage error is 1% and the final element errors are shown in brackets.

7.2. Optimum Electrode Positioning

A study to examine the effects on reconstruction using different electrode configurations is done through spectral analysis. How spectral analysis can help in determining the most reliable reconstruction is shown and then applied to different electrode configurations.

7.2.1. Spectral Analysis

This section examines the importance of the set of Jacobian singular values on the solution accuracy. This set of singular values is commonly known as the spectrum. The well known Singular Value Decomposition (SVD) will find the spectrum and can also be used to solve the least squares problem. It is a time consuming solution method and the Cholesky decomposition is the preferred choice for the actual equation solution. However, the SVD demonstrates the effects of the spectrum on the solution accuracy.

The importance of the spectrum on the accuracy in the reconstruction is most easily demonstrated solving the least squares problem with Levenberg-Marquardt regularisation, as shown in equation (5.20). The SVD solution is described in Golub[89],

- Compute the SVD of the Jacobian matrix, $\mathbf{J} = \mathbf{U}\Sigma\mathbf{Z}^{\mathrm{T}}$, where $\mathbf{U}$ and $\mathbf{Z}$ are orthogonal matrices, whose columns, $\mathbf{U}_i \& \mathbf{Z}_i$, are respectively the left and right singular vectors. The diagonal matrix Σ contains the set of positive singular values, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m$ or the spectrum. The number of elements in the FEM mesh is m and for all m singular values to be greater than zero, the projection angles must provide enough independent information.
- Form the vector, $\mathbf{b} = \mathbf{U}^{\mathrm{T}}\left[\mathbf{f} - \mathbf{v}_{\mathrm{o}}\right]$.
- The new update at iteration k, $\Delta\rho^k$ is solved using

$$\Delta\rho^k = \sum_{i=1}^{m} w_i \cdot \mathbf{Z}_i \tag{7.5}$$

where the regularisation parameter is denoted by λ and w_i is a scalar weighting factor:

$$w_i = \frac{\sigma_i}{\sigma_i^2 + \lambda}\mathbf{b}_i \tag{7.6}$$

Equation (7.5) shows that the resistivity update is the weighted sum of the right singular vectors. Our interest lies in the weighting factor. With no regularisation, that is $\lambda = 0$, when any of $\sigma_i \to 0$ then $w_i \to \infty$. The summation then includes a very heavy weighting of the right singular vector. However, both the right singular vectors and $\mathbf{b}_i$ include practical system errors. The errors are magnified by the weighting factor, swamping the solution with the errors. By introducing the regularisation parameter, the weighting factor is controlled to some degree so that large errors are not introduced into the summation. However, the larger the regularisation parameter, the more the true information becomes masked. The choice of the regularisation parameter has already been discussed to balance these effects for optimum solution. It becomes clear that the larger the singular values, the less the effects of noise and the less regularisation that needs to be introduced.

The spectrum is said to indicate the degree of conditioning in the problem. Well conditioned problems will have an even spread of singular values. Ill-conditioned problems include small singular values in the spectrum. The condition number is important in studying the conditioning as it is equal to the ratio of the largest to the smallest singular value, see Appendix A2. However, a graphical display of the full spectrum gives a better indication of the conditioning in the problem and the likely success in the reconstruction. The full spectrum will be graphed in later sections when examining different electrode configurations.

7.2.2. Initial Study

An initial study was undertaken which examined the effects of using internal electrodes. This study used a small square 25 element mesh and the unregularised Gauss-Newton method for reconstruction. Two different electrode configurations were compared. The Edge electrode configuration uses 20 electrodes around the periphery of the mat. The Internal configuration uses the same 20 electrodes, except 4 of them are replaced by the 4 most central mesh nodes. Some of the projection angles use the central electrodes for current injection. The results are shown in Figures 7.3 to Figure 7.5 for a load spread throughout the centre of the mat with a low voltage measurement error. Figure 7.8 compares the quality of reconstruction for a single centrally perturbed element in the presence of higher noise.

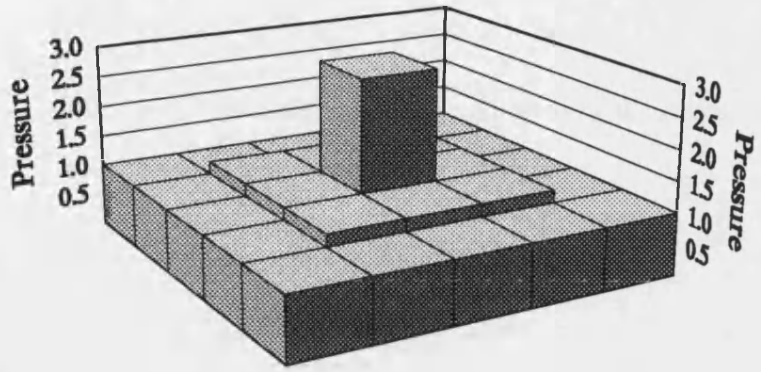


Figure 7.3 The 25 element test profile

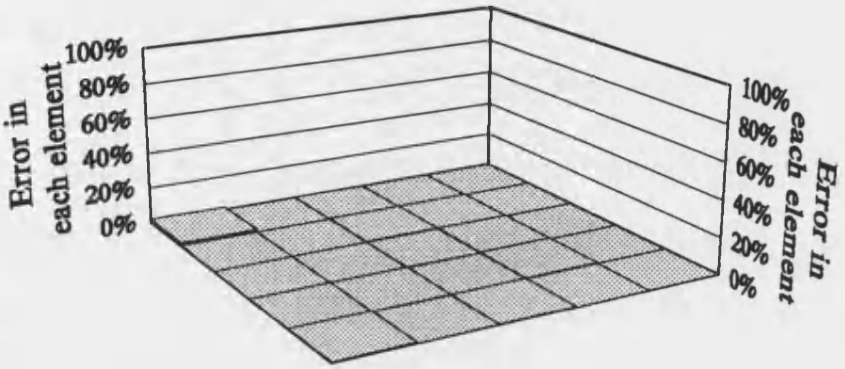


Figure 7.4 The 25 element test reconstruction errors using the Internal electrode configuration. Voltages have a 0.3% measurement error.

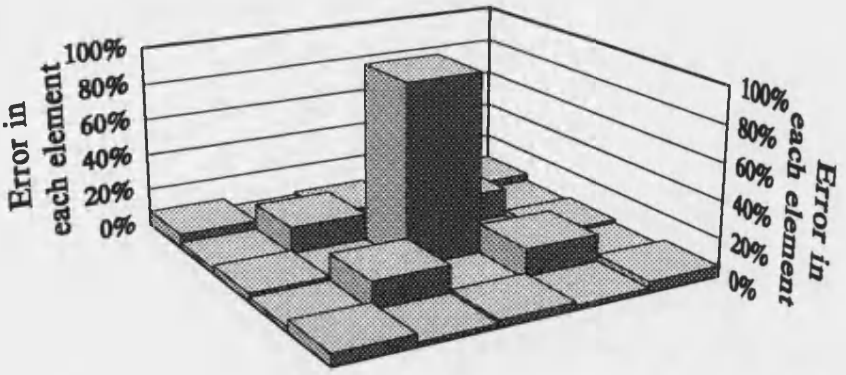


Figure 7.5 The 25 element test reconstruction errors using the Edge electrode configuration. Voltages have a 0.3% measurement error.

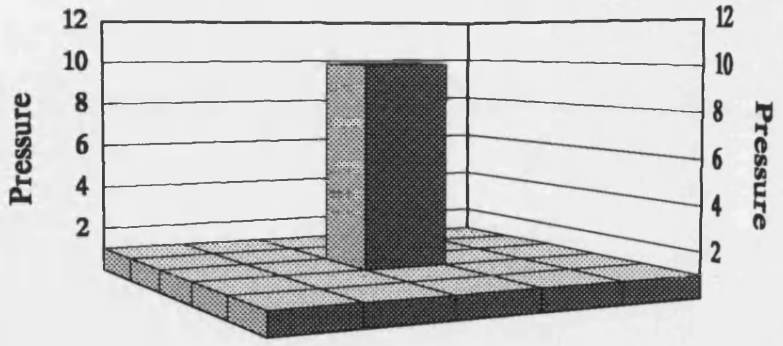


Figure 7.6 The single centrally perturbed element test profile.

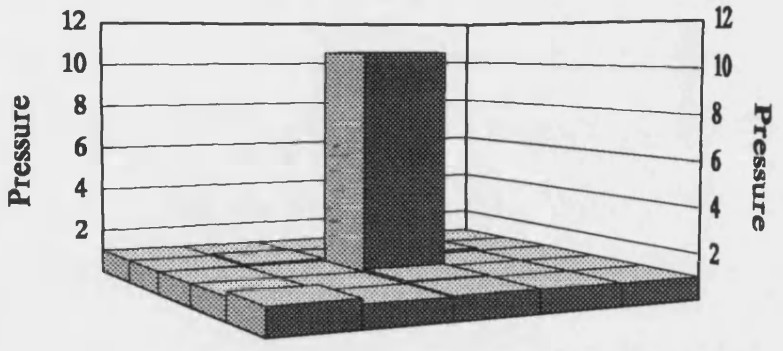


Figure 7.7 Reconstruction for the single centrally perturbed element test using the Internal electrode configuration. Voltage measurement error is 3%.

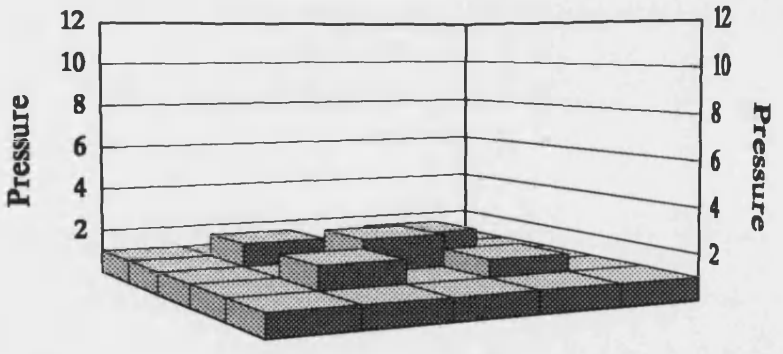


Figure 7.8 Reconstruction for the single centrally perturbed element test using the Edge electrode configuration. Voltage measurement error is 3%.

It is quite clear that by placing a mere 4 electrodes in the centre of the mat, the reconstructions improve dramatically. Experimentation with the position of the pressure distribution showed that the improvements were large, but not as impressive for distributions near the edge of the mat. The test distributions shown here are for pressure distributions based in the centre of the mat. Unlike the Internal configuration, many of the test distributions failed to converge using the Edge configuration and single digit voltage measurement errors. In this case, the error destabilised the Gauss-Newton method, but can be stabilised through regularisation. A complete study using practical pressure distributions and regularisation on larger meshes follows in the next section.

Fulton[92] showed similar results with the larger square 64 element mesh and correlated the improved accuracy with an improved matrix condition number. Results presented by Moskowitz[93] at the same conference, used a limited number of invasive electrodes in the application of EIT to medical thermal imaging. A sharper, more accurate reconstruction of the change in conductivity with temperature of a heated solid phantom in a conducting liquid was obtained.

Pilkington[94] examined the distinguishability in impedance imaging problems when a single intracavity reference electrode is used in addition to those on the surface of the body. The intracavity electrode was intended to be placed on the heart or lungs. A two-dimensional circular model examining single centrally placed targets of different radii produced theoretical distinguishability calculations. These showed that the distinguishability increased by a factor greater than 6 for targets less than 10% in radius. Although no results using reconstructions have consequently been produced, the discussion on optimal currents in Section 7.1.2 shows that distinguishability improvements should lead to better reconstructions. The results in the following section compare different electrode configurations examining the reconstructed pressure distributions and their associated distinguishabilities.

7.2.3. Electrode Positioning Results

The variables which have an effect on the reconstruction success are outlined at the beginning of Chapter 7. The experiments in this section compare different electrode positions by keeping the other variables constant. The aim is to find the best set of electrode positions for use in an EIT pressure measuring system.

7.2.3.1. Set-up

The Gauss-Newton method with second finite differences regularisation has been shown to be the best for an EIT pressure sensitive problem and is the solution method used from here on. Similarly, the optimal current patterns, described earlier, are used in order to reduce the effect of differing current patterns on reconstruction. As mentioned, the currents are optimal for the first iteration as this will be the most likely practical approach to using optimal currents. The practical environment simulated, is the same as the one used in Section 6.3.2, that is, one with a 1% voltage error measurement. The experiments were carried out using meshes that provide reasonably good spatial resolution, namely, the 100, 256 and 400 element grids. The greater the number of electrodes used, the better the conditioning of the problem. Any electrode configuration used in a comparison thus uses the same number of electrodes; either 36 or 37.

7.2.3.2. Optimal Electrode Positions

Knight[1] examined the sensitivity of each electrode to resistivity perturbations in each element. He concluded that the sensitivity of an electrode is a function of the distance of the electrode from the conductance change that it attempts to detect. Paulson[50] also notes that the sensitivity drops off dramatically as perturbations are made further from the surface in his imaging of the body using surface electrodes. In an EIT pressure scanner system, the electrodes can be placed anywhere on the mat. All electrodes connected to the mat are point electrodes, including those connected to the interior of the mat. The wires leading to the internal electrodes must, of course, be electrically isolated from the mat. The wires must also be thin and flat so that they do not introduce a disturbance to the pressure distribution being profiled.

The ideal or optimal electrode position configuration would then be one where there is an electrode below any load on the mat. This can be verified by setting up test electrode configurations. Flexibility in placing the electrodes anywhere on the domain is taken advantage of and the Concentrated electrode pattern shown in Figure 7.9 is used. Test pressure profiles needed are those directly on top of the concentrated electrodes, as well as those varying in distance from these electrodes. Such a set of distributions and the resulting reconstructions are shown in Figures 7.10 to 7.12. Each of the test profiles are identical, merely moved to different positions.

The reconstructions show quite distinctly that the more electrodes there are in the vicinity of the test profile, the better is the reconstruction. Profile TOPLEFT is directly over the electrodes and produces the most accurate reconstruction. The test OFFSET has half the distribution over the electrodes and produces the next best reconstruction. The BOTTOMRIGHT distribution is furthest away from the electrodes and looks the least like the actual test pattern. The element errors graphed in Figure 7.13 quantify this trend.

The distinguishabilities measured at the first iteration are graphed in Figure 7.14. The graphs show that there is a lower error in the reconstruction when the distinguishability is higher. The larger distinguishability leads to better reconstructions as there is a larger signal to noise ratio in the voltage measurements. These results confirm that the distinguishability is higher for test distributions in close vicinity of electrodes.

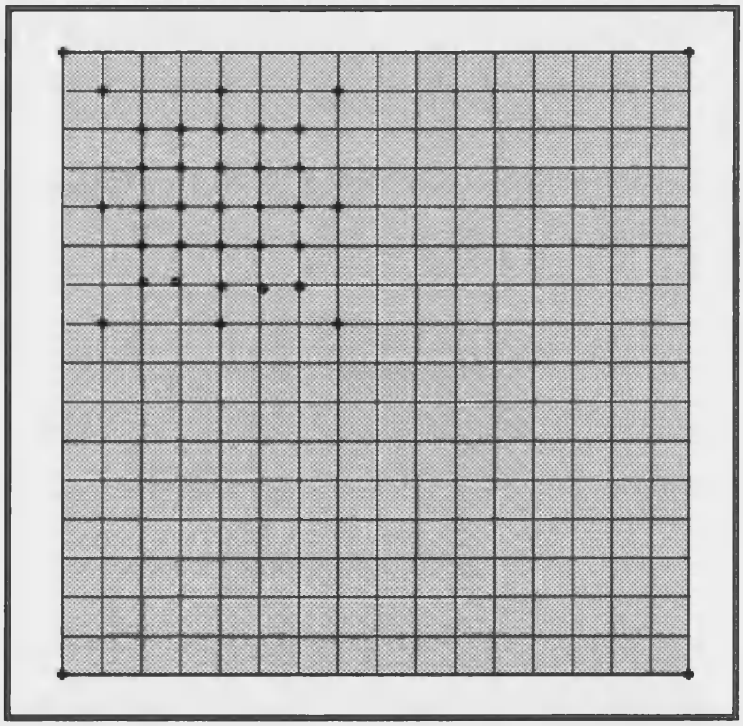


Figure 7.9 The 37 Concentrated electrode configuration. The mesh used has 256 elements.

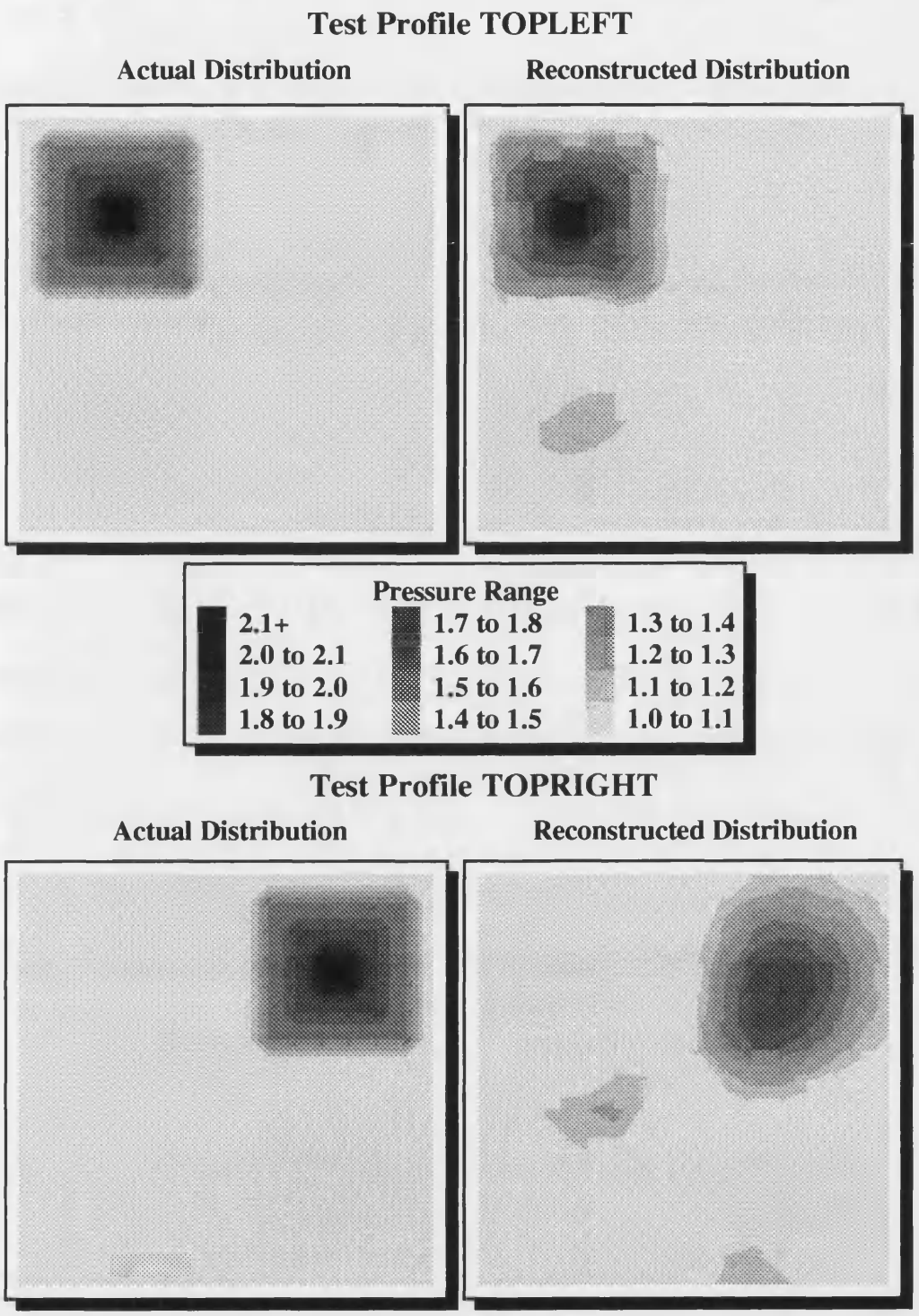


Figure 7.10 Test distributions TOPLEFT and TOPRIGHT with the reconstructed distributions using the Concentrated electrode configuration.

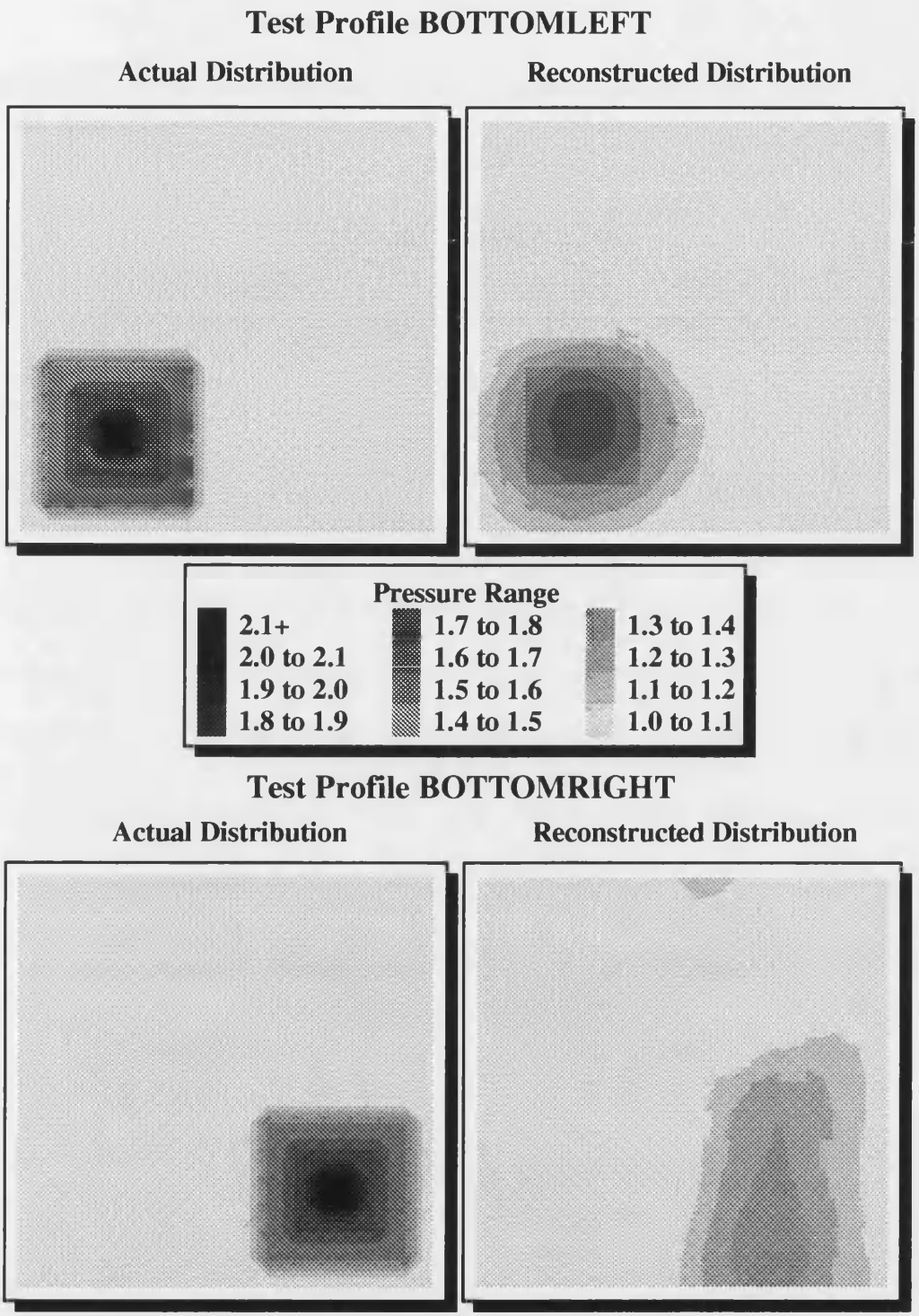


Figure 7.11 Test distributions BOTTOMLEFT and BOTTOMRIGHT with the reconstructed distributions using the Concentrated electrode configuration.

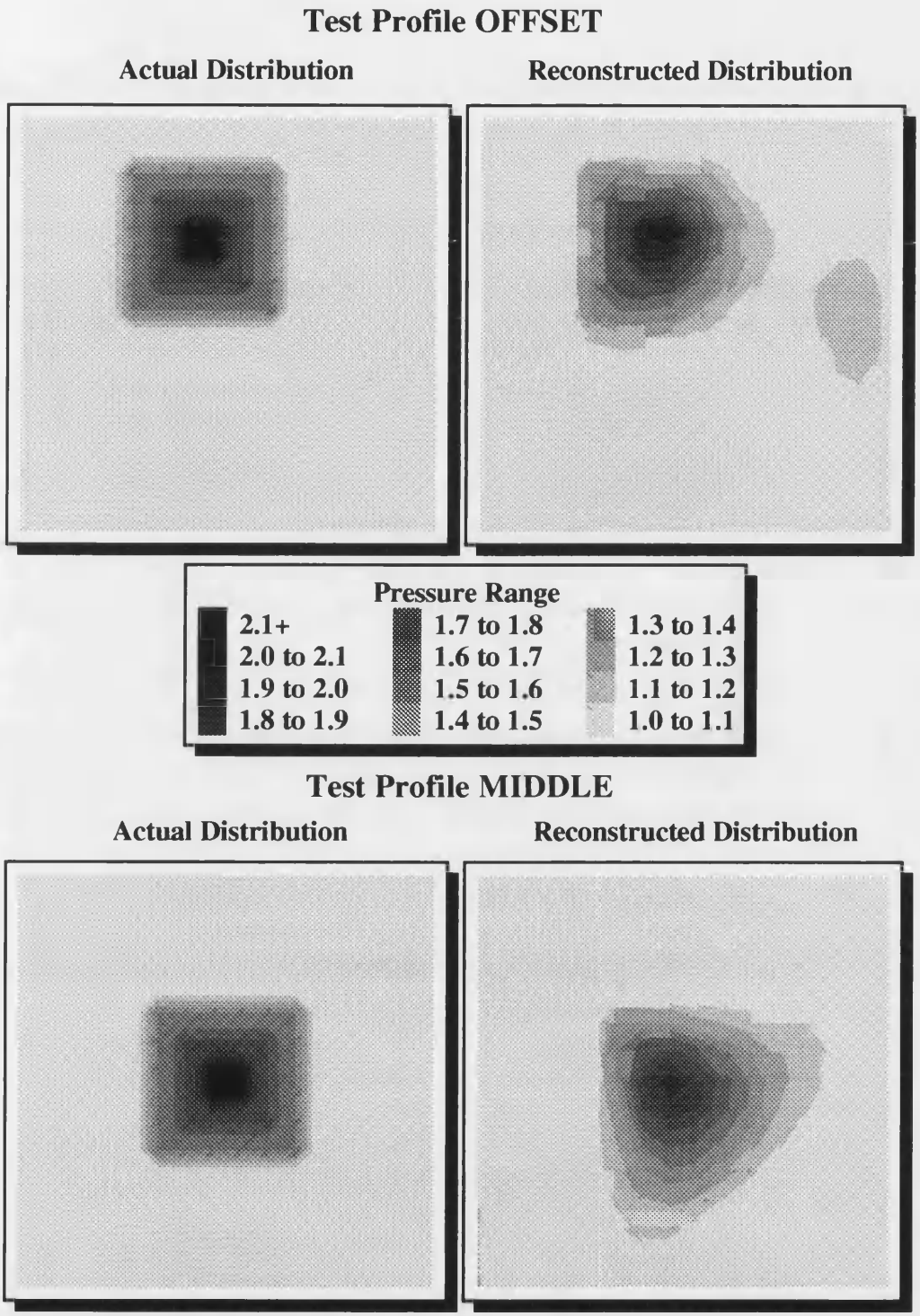


Figure 7.12 Test distributions CENTRE and OFFSET with the reconstructed distributions using the Concentrated electrode configuration.

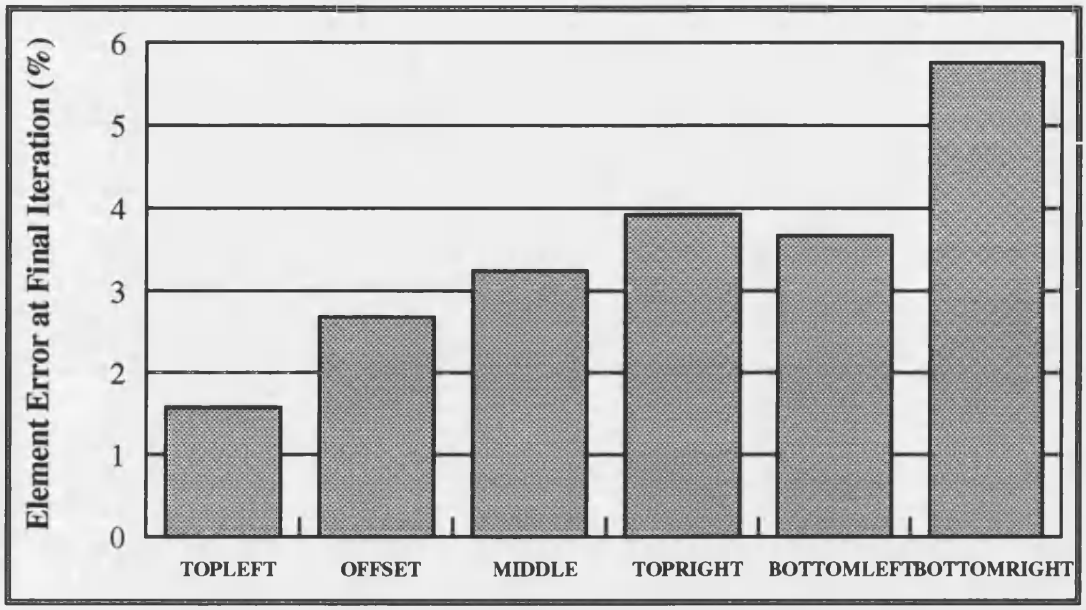


Figure 7.13 Reconstruction errors for 6 test distributions on the 256 element grid using the Concentrated electrode configuration.

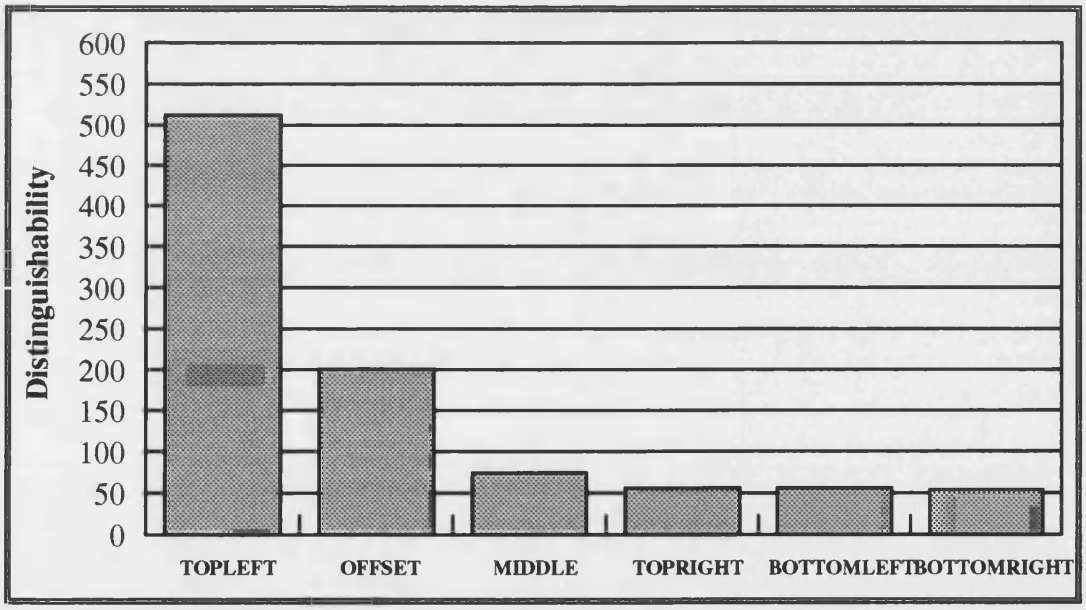


Figure 7.14 Distinguishabilities for 6 test distributions on the 256 element grid using the Concentrated electrode configuration.

7.2.3.3. Practical Electrode Positions

The importance of having the electrodes in close proximity of the pressure distribution has been established. However, in any practical system, this is not possible as the pressure distribution is constantly changing. In standard EIT applications, the electrodes are distributed around the edge of the domain, however, the region of interest is usually towards the centre of the domain. The most practical approach for the average pressure profile, would be to distribute the electrodes uniformly throughout the mat. This section looks at the improvements this type of electrode configuration has over traditional EIT configurations of edge electrodes. The underlying reasons for improvements are examined in the following section.

The Edge electrode configuration used is shown in Figure 7.15 for test patterns on the 256 element mesh and in Figure 7.17 for the 100 and 400 element meshes. The proposed practical electrode positions using the entire mat area are shown in Figures 7.16 and 7.18. These are named the Internal electrode configurations. Note that for any particular mesh, the number of electrodes used is the same for both configurations. The 100, 256 and 400 element mesh test distributions of Section 6.3.1 are the main test patterns utilised. The resulting reconstructions are shown in Figures 7.19 to 7.32 with a voltage measurement error of 1%.

Studying the reconstructions, it can be seen that the Internal configuration always results in a superior reconstruction. The element errors indicate this, even when some of the reconstructions 'look' fairly similar. The narrowest gap in reconstruction errors is with the 400 element FOOT distribution. In this case, the Edge reconstructed foot is perhaps a better looking foot, but on closer inspection, the reconstruction with the Internal configuration is closer to the test shape. The Edge configuration tends to smooth the distribution and this is most clear in the BLOCK profiles. There is also more sign of the big toe in the 256 mesh FOOT Internal reconstruction. The Internal configuration always gives greater accuracy in the central areas. This can be seen, for example, in the area between the buttocks in the WHEELCHAIR profiles. The best verification of this is the reconstruction of Figure 7.27 which uses a tiny central test profile; TINYCENTRE. The Internal test produces an image almost identical to the actual distribution, whereas the Edge test is almost a total failure.

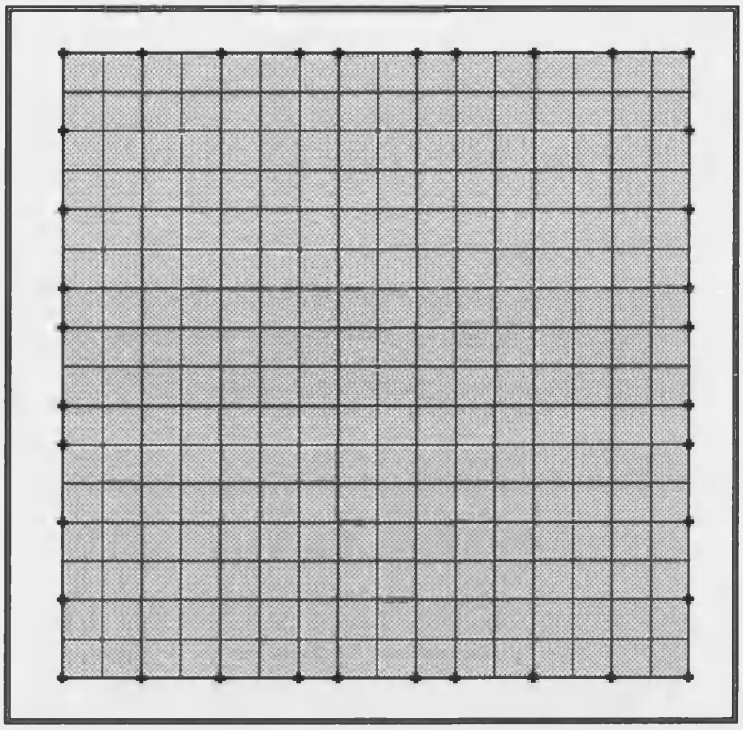


Figure 7.15 The 37 Edge electrode configuration for grids with 256 elements

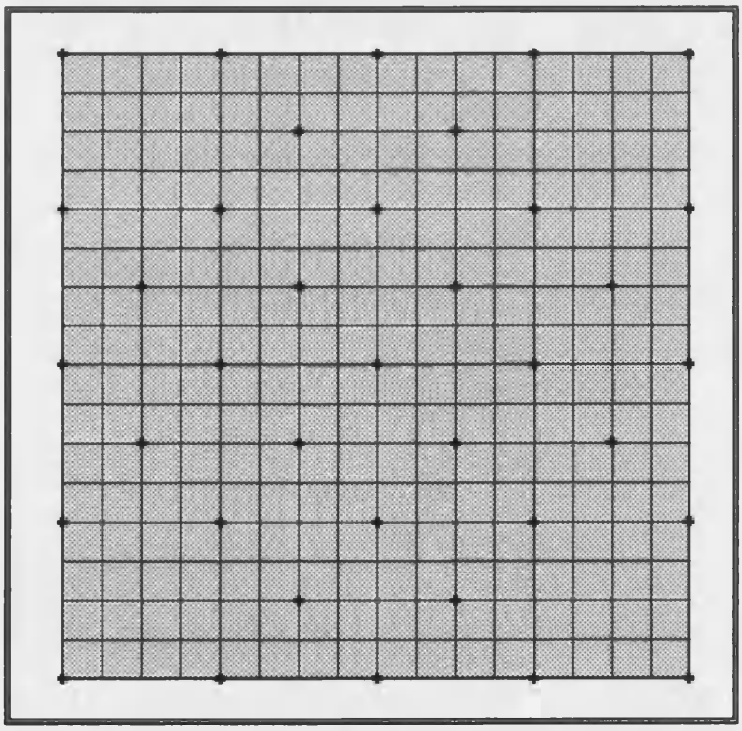


Figure 7.16 The 37 Internal electrode configuration for grids with 256 elements

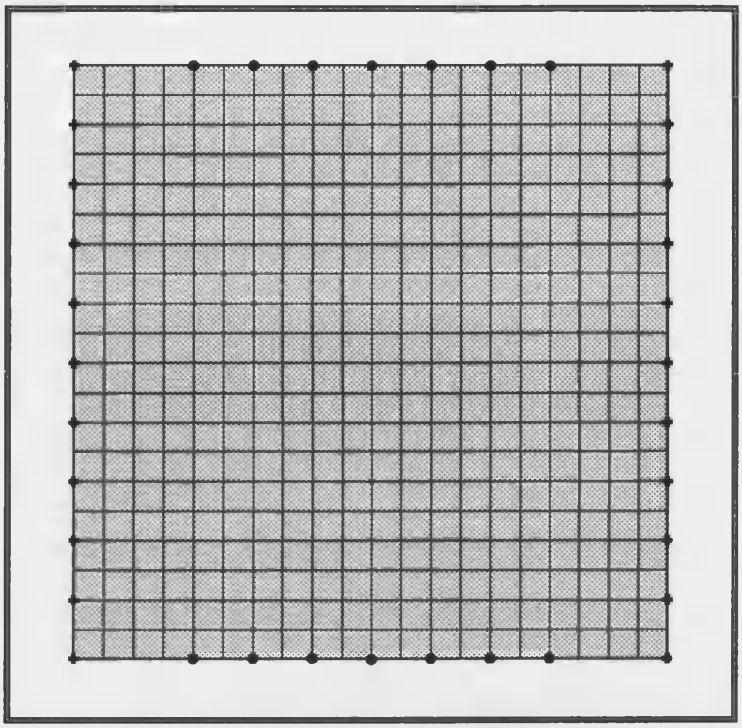


Figure 7.17 The 36 Edge electrode configuration for grids with either 100 or 400 elements

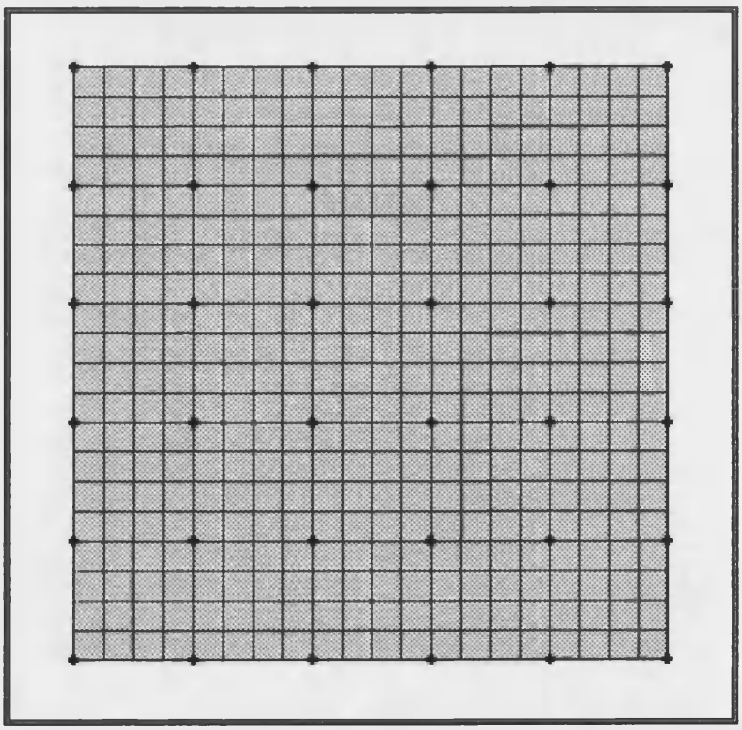


Figure 7.18 The Internal 36 electrode configuration for grids with either 100 or 400 elements

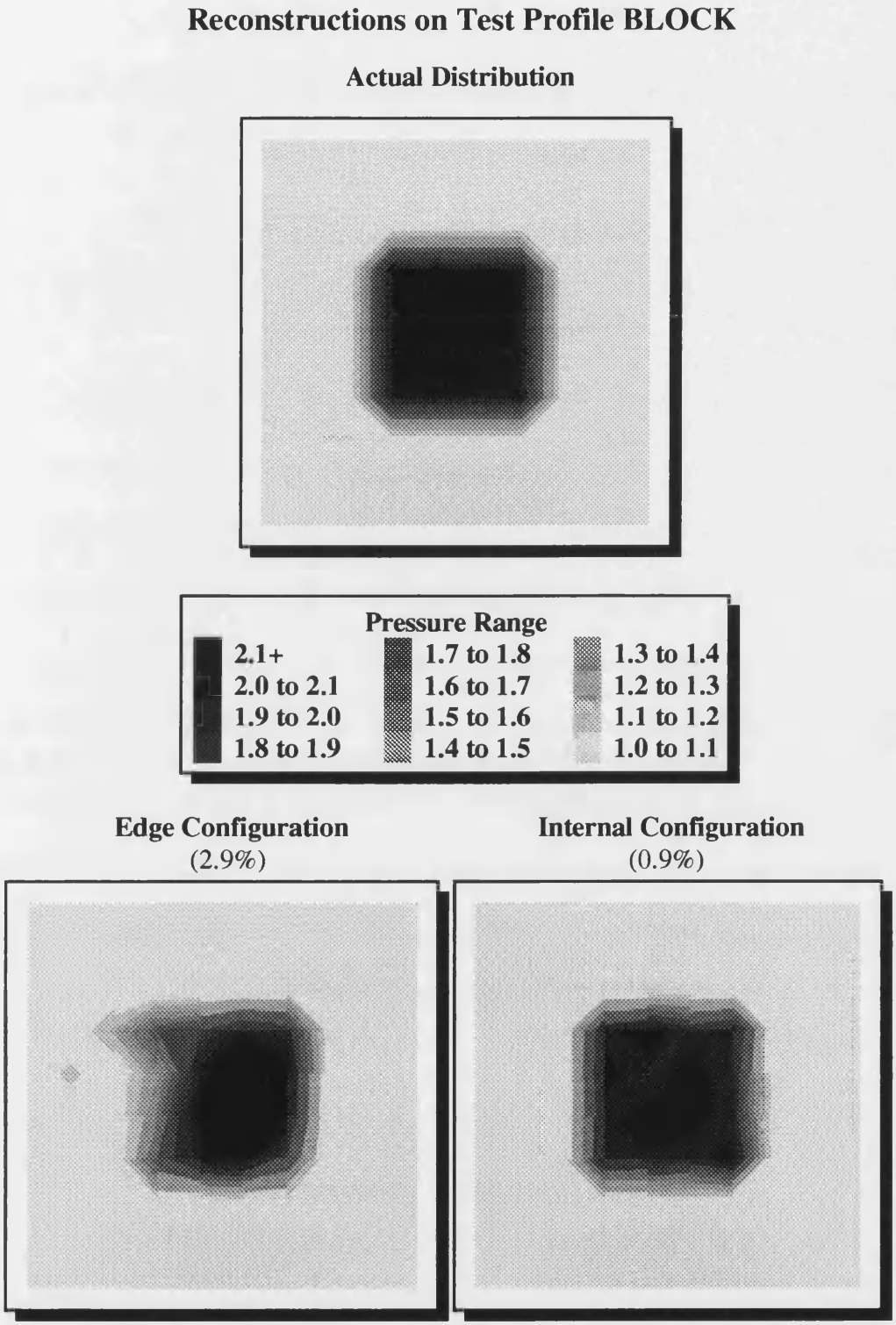


Figure 7.19 Reconstructions on the 100 element mesh using test distribution BLOCK comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

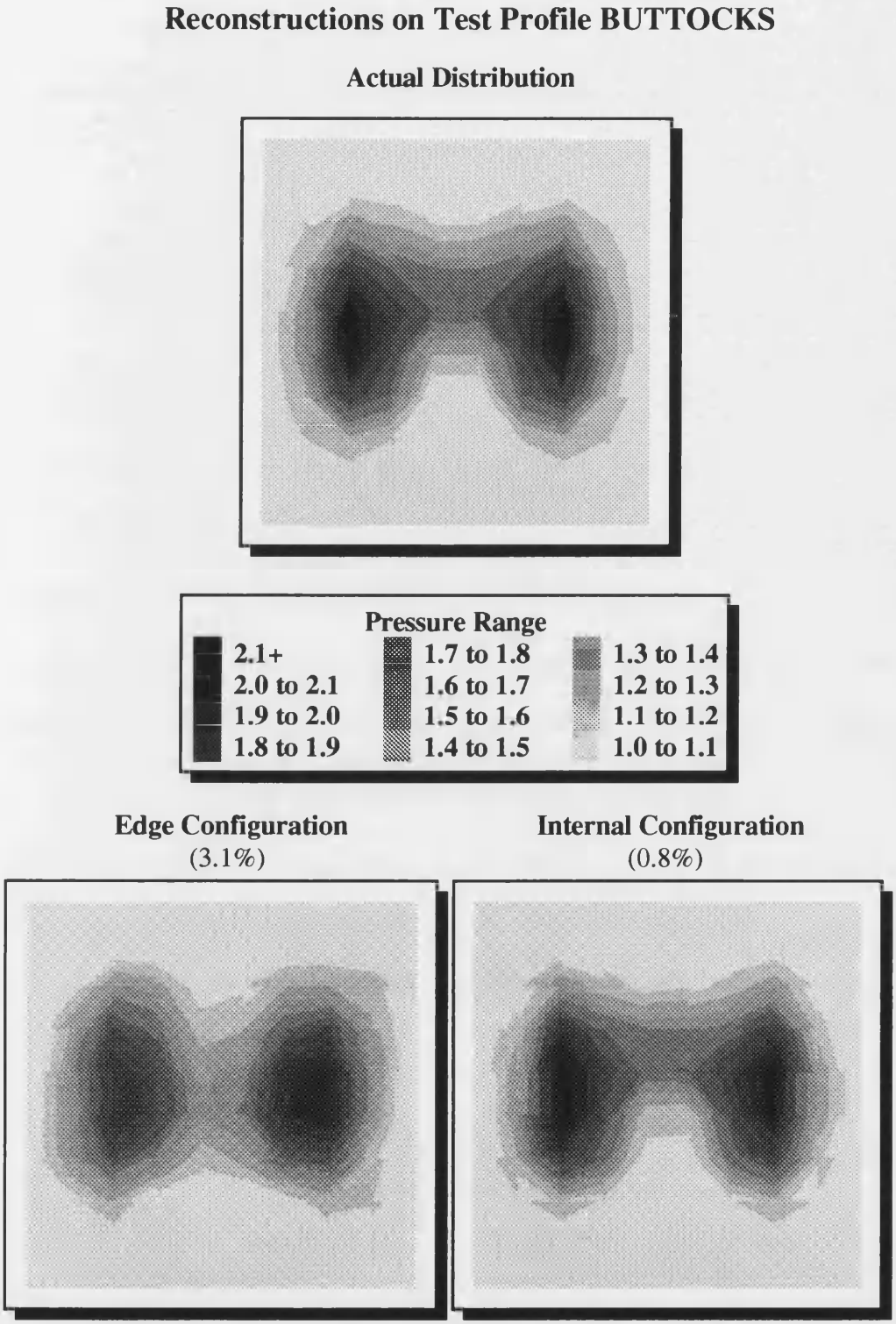


Figure 7.20 Reconstructions on the 100 element mesh using test distribution BUTTOCKS comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

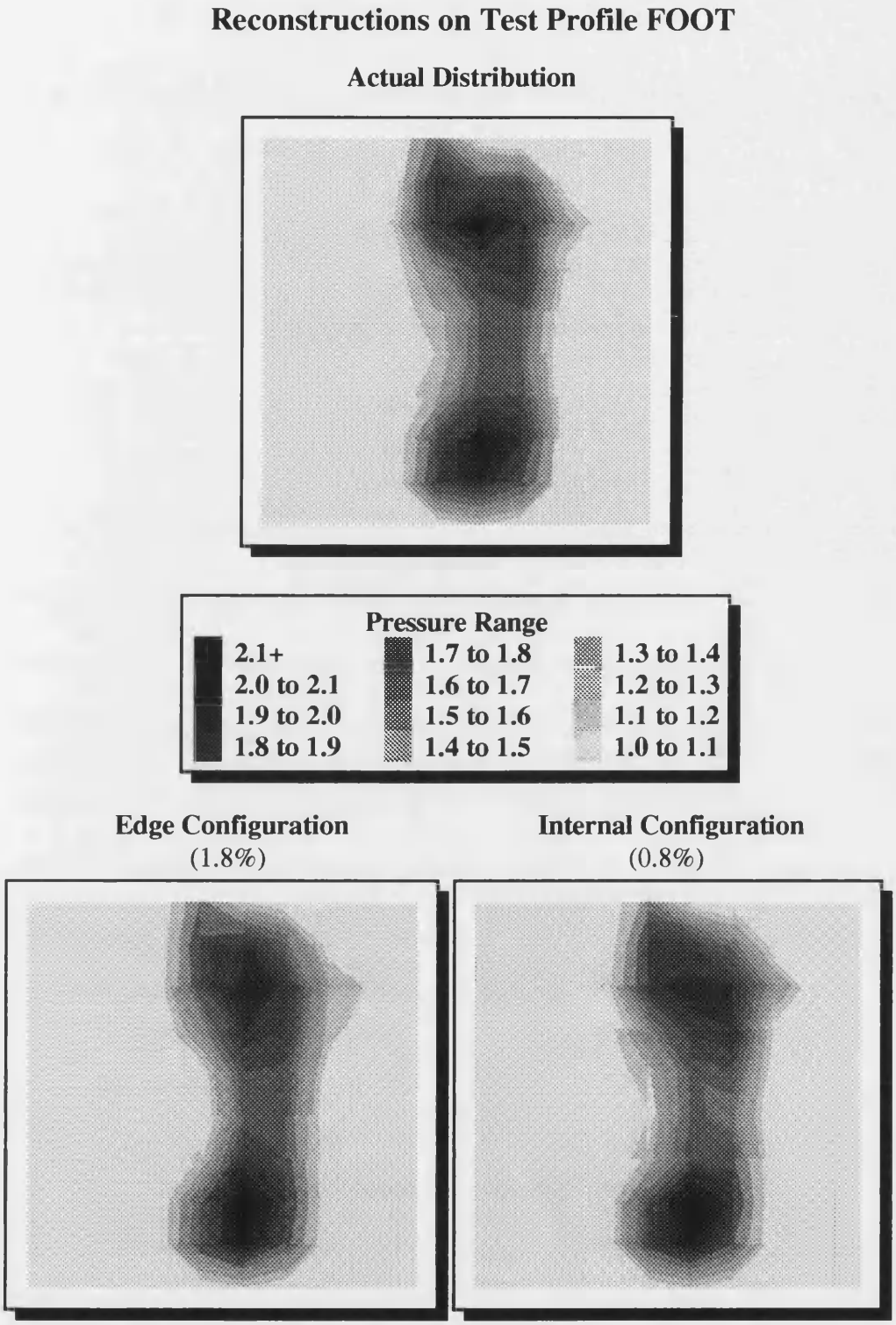


Figure 7.21 Reconstructions on the 100 element mesh using test distribution FOOT comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

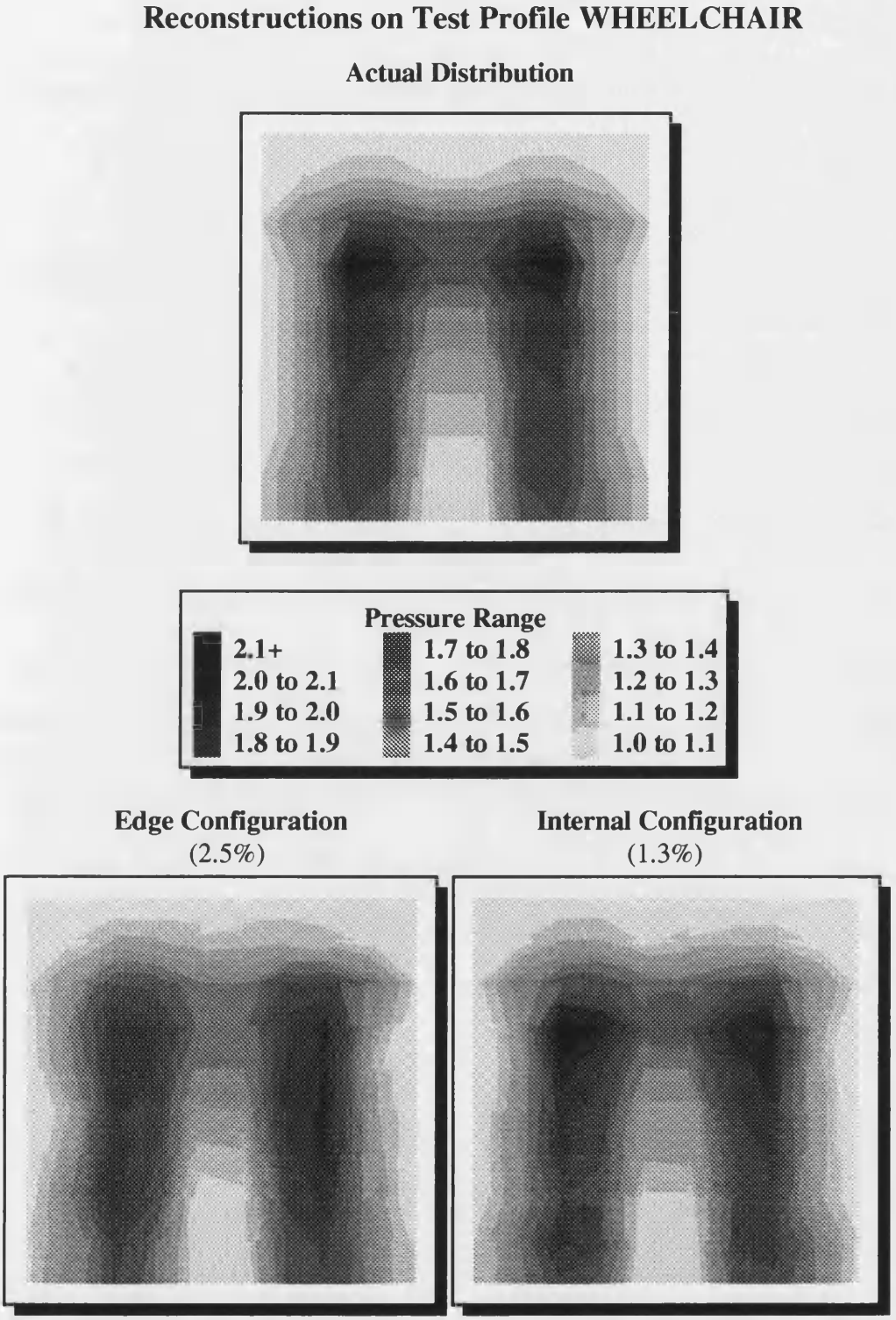


Figure 7.22 Reconstructions on the 100 element mesh using test distribution WHEELCHAIR comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

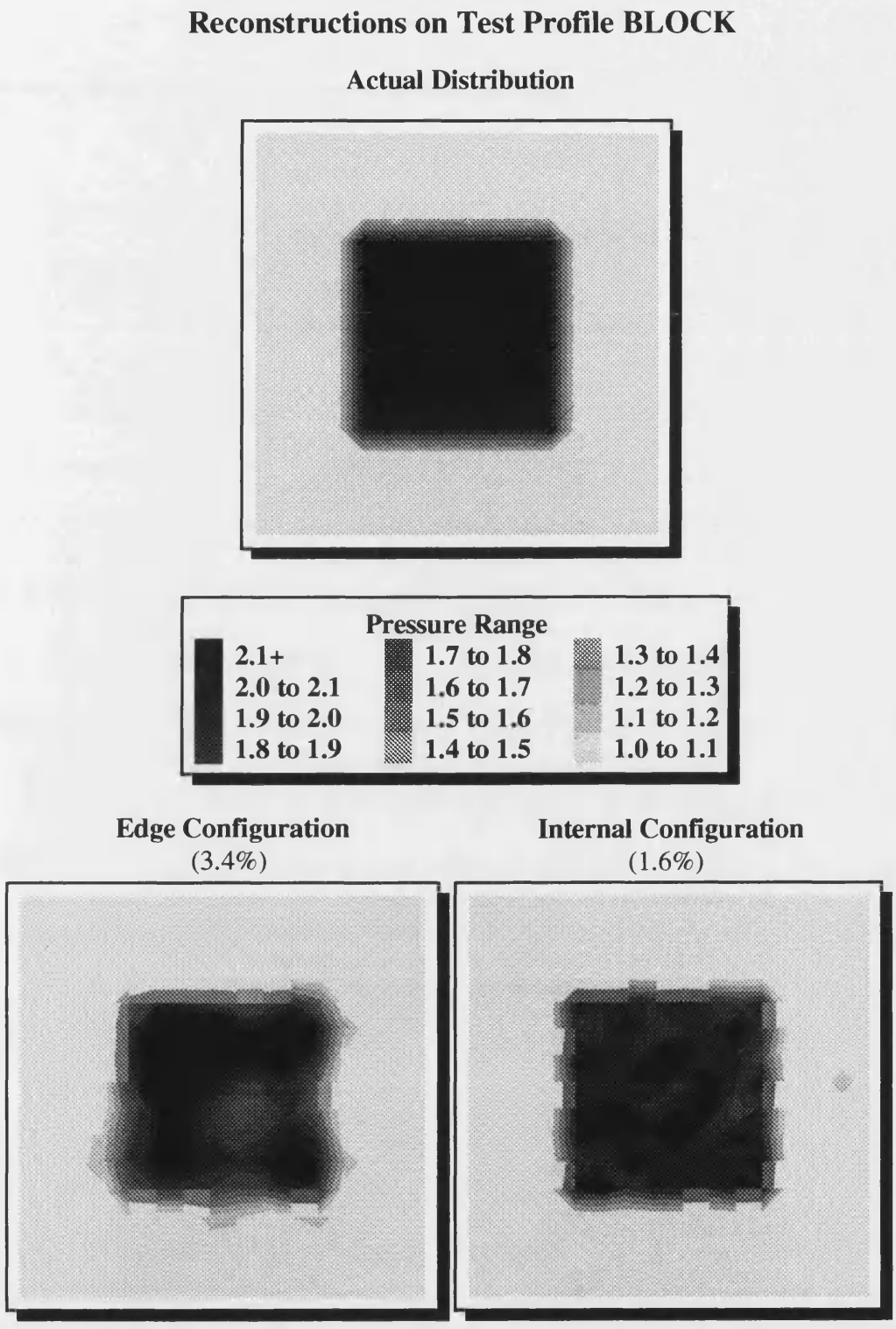


Figure 7.23 Reconstructions on the 256 element mesh using test distribution BLOCK comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

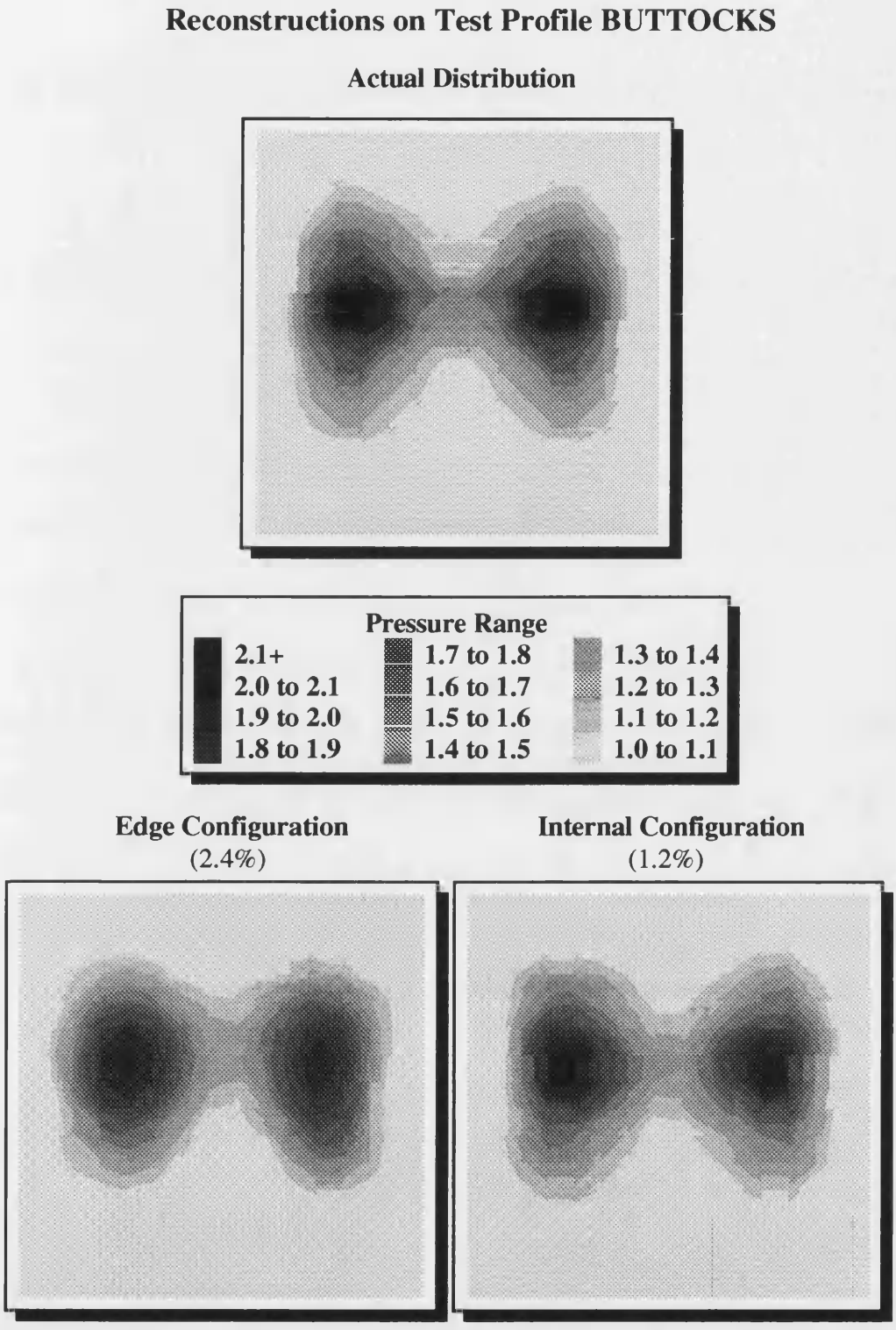


Figure 7.24 Reconstructions on the 256 element mesh using test distribution BUTTOCKS comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

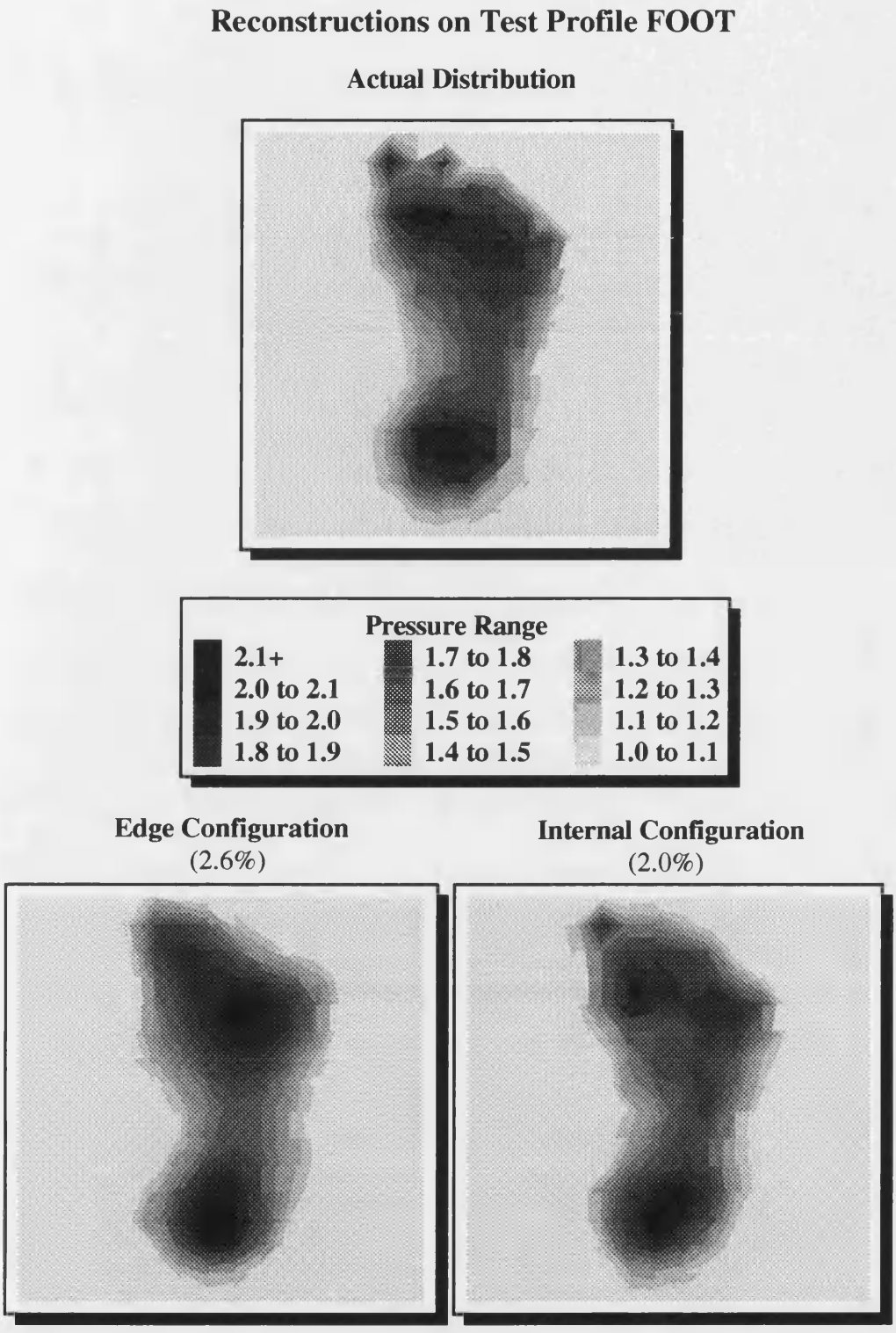


Figure 7.25 Reconstructions on the 256 element mesh using test distribution FOOT comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

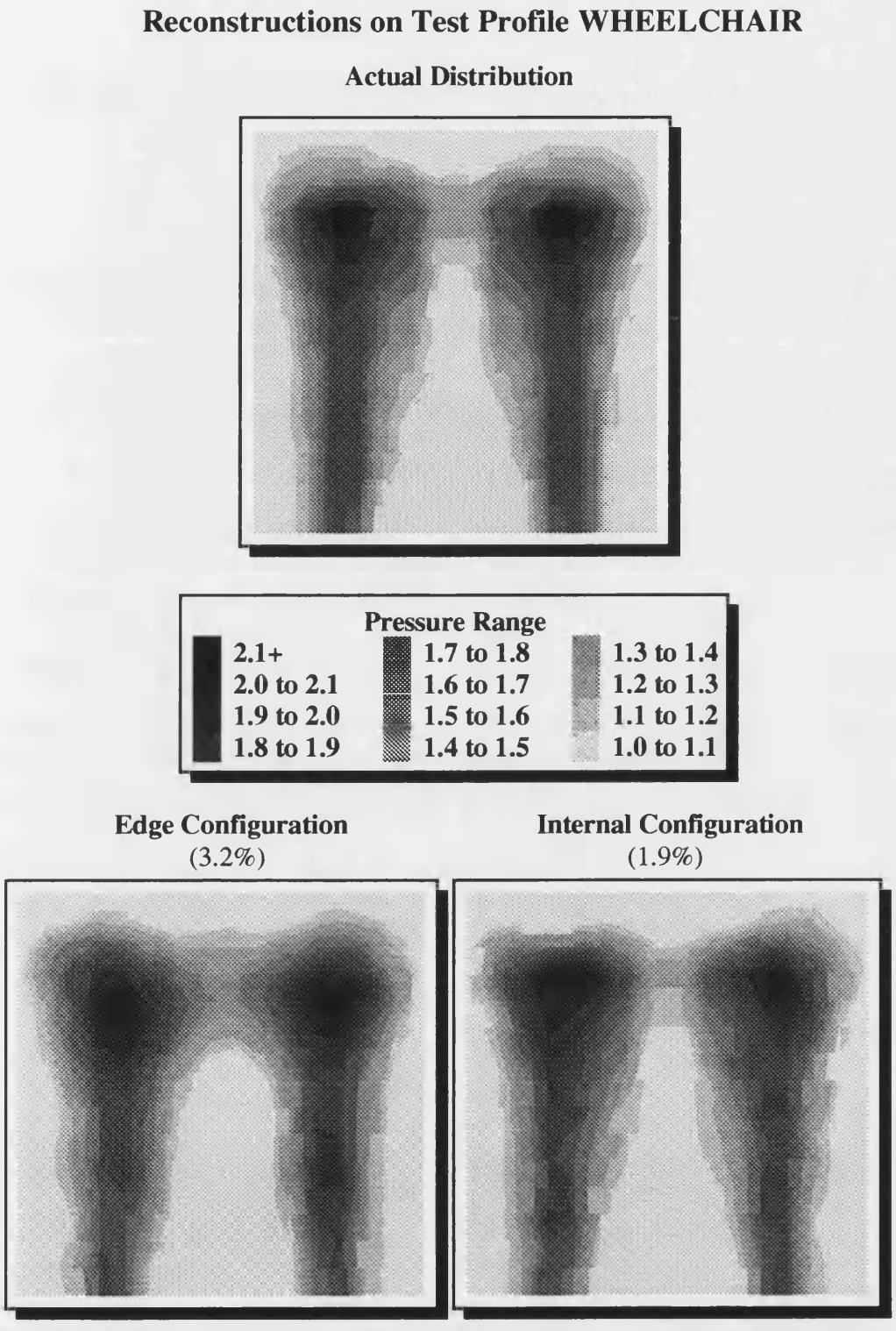


Figure 7.26 Reconstructions on the 256 element mesh using test distribution WHEELCHAIR comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

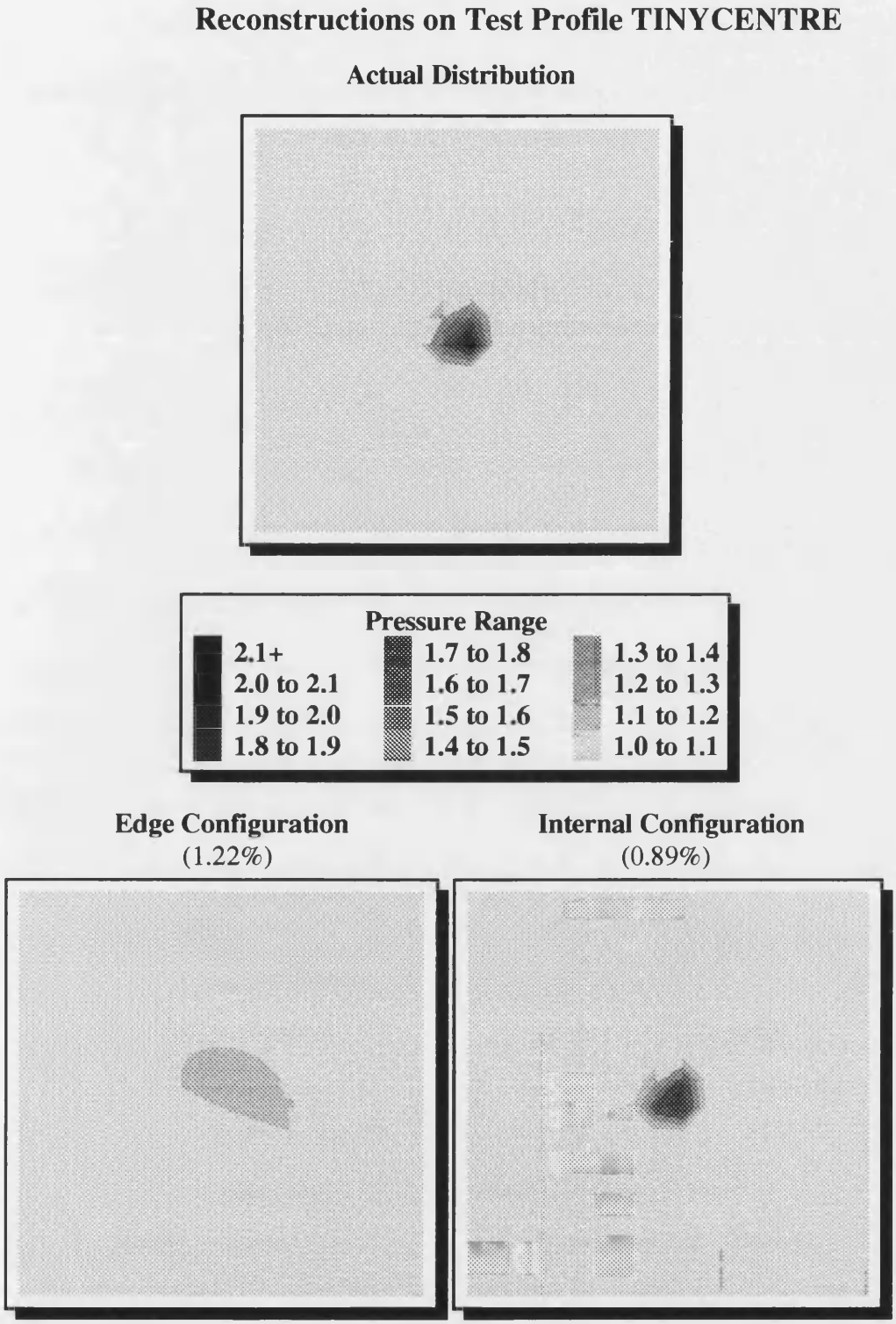


Figure 7.27 Reconstructions on the 256 element mesh using test distribution TINYCENTRE comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

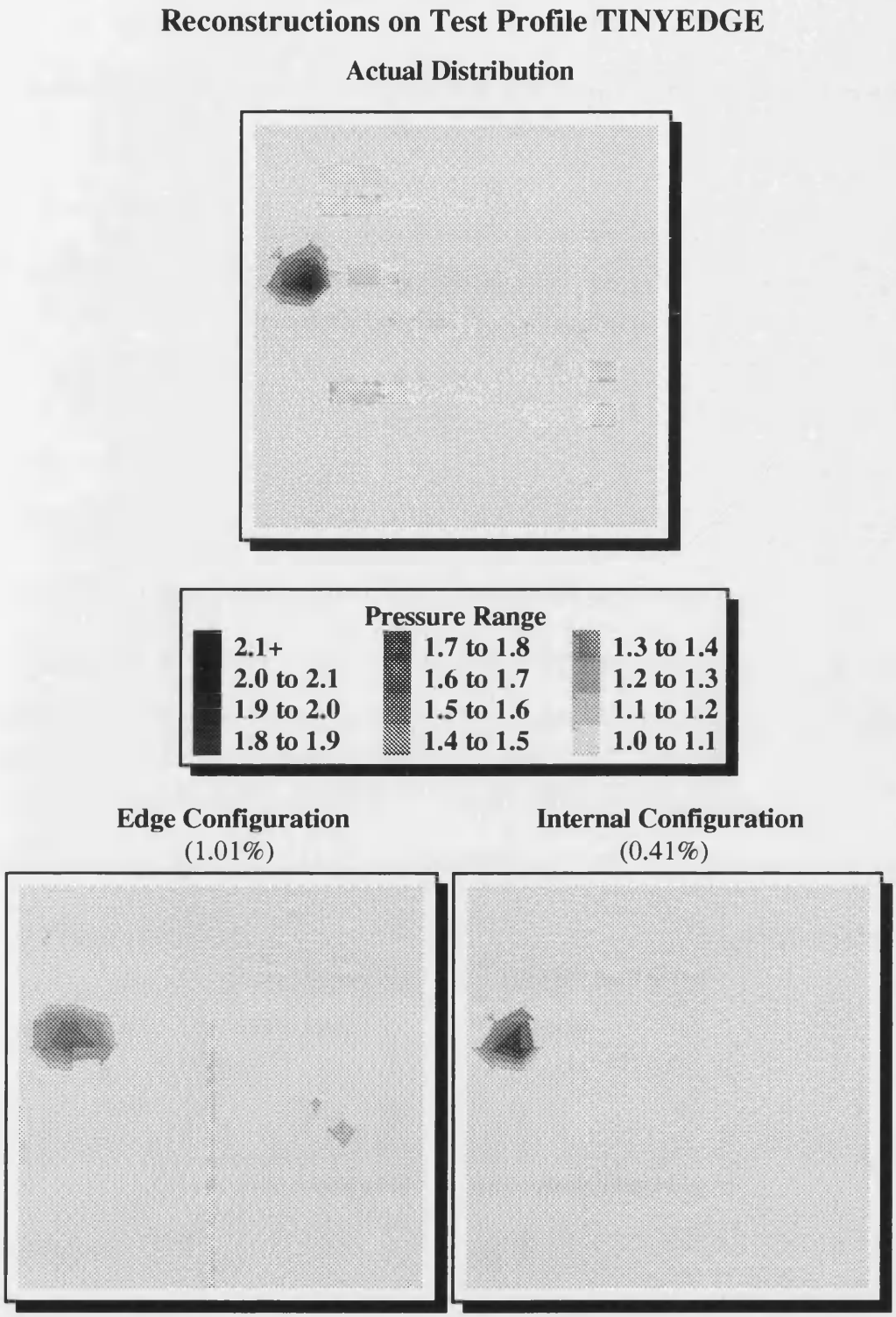


Figure 7.28 Reconstructions on the 256 element mesh using test distribution TINYEDGE comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

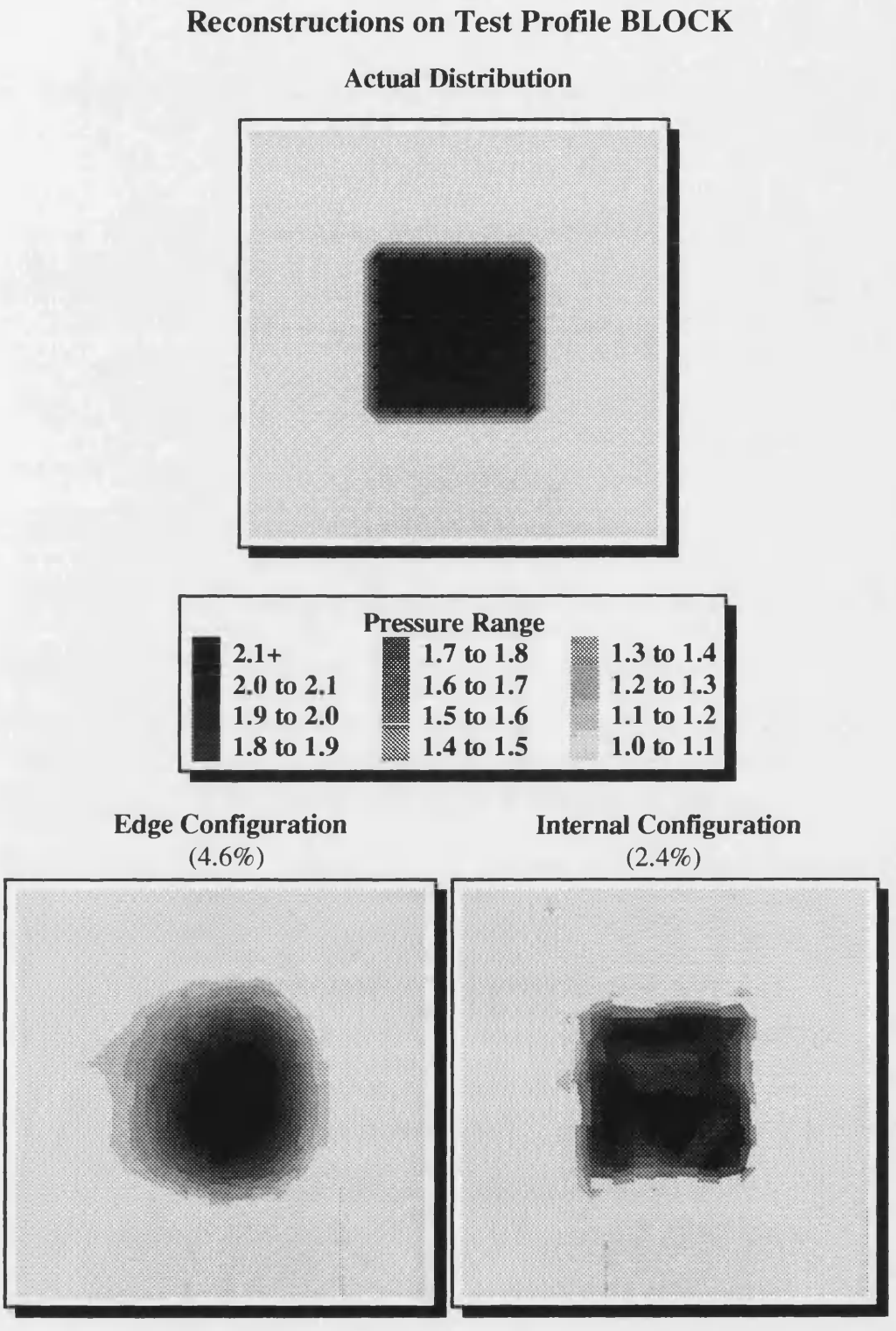


Figure 7.29 Reconstructions on the 400 element mesh using test distribution BLOCK comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

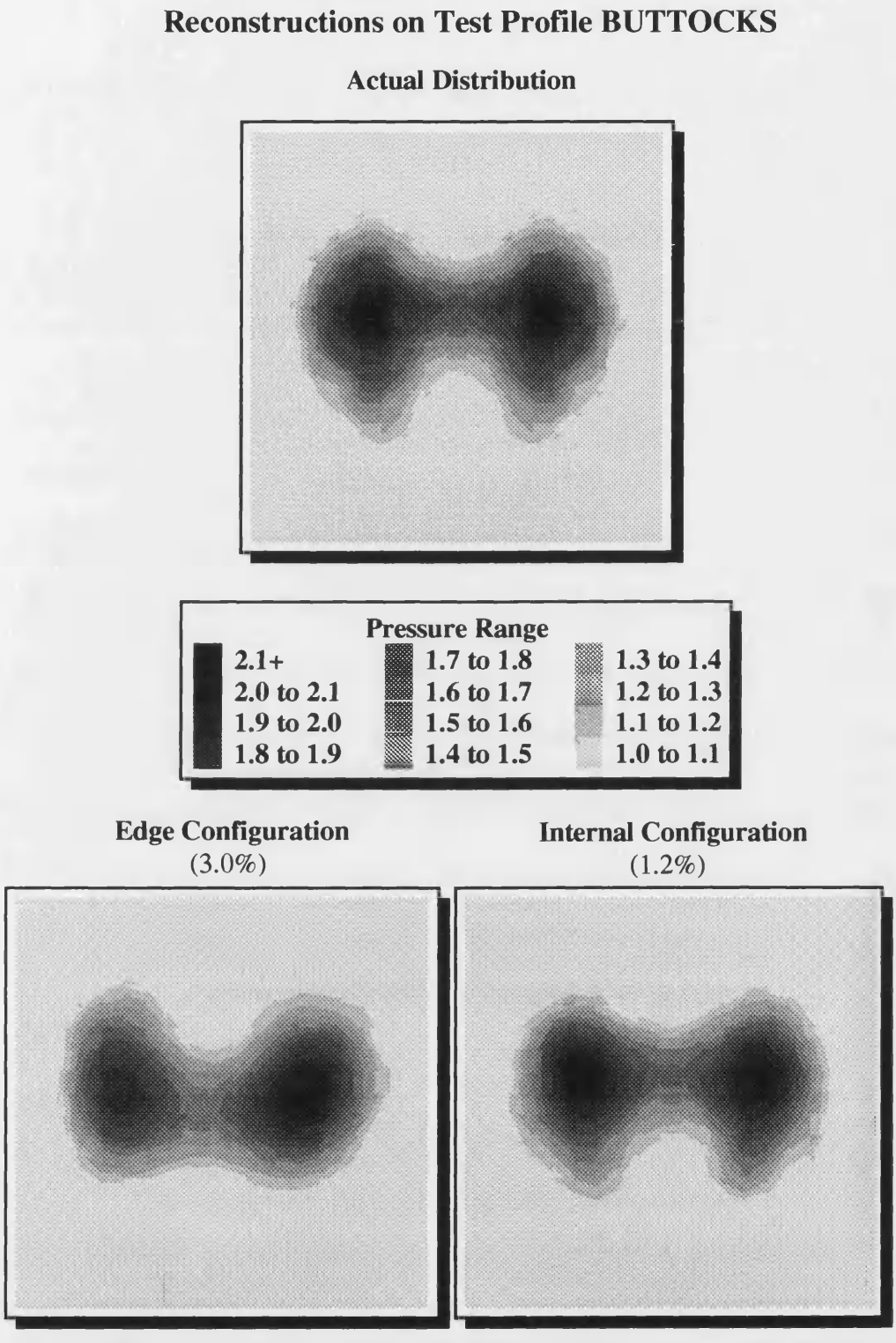


Figure 7.30 Reconstructions on the 400 element mesh using test distribution BUTTOCKS comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

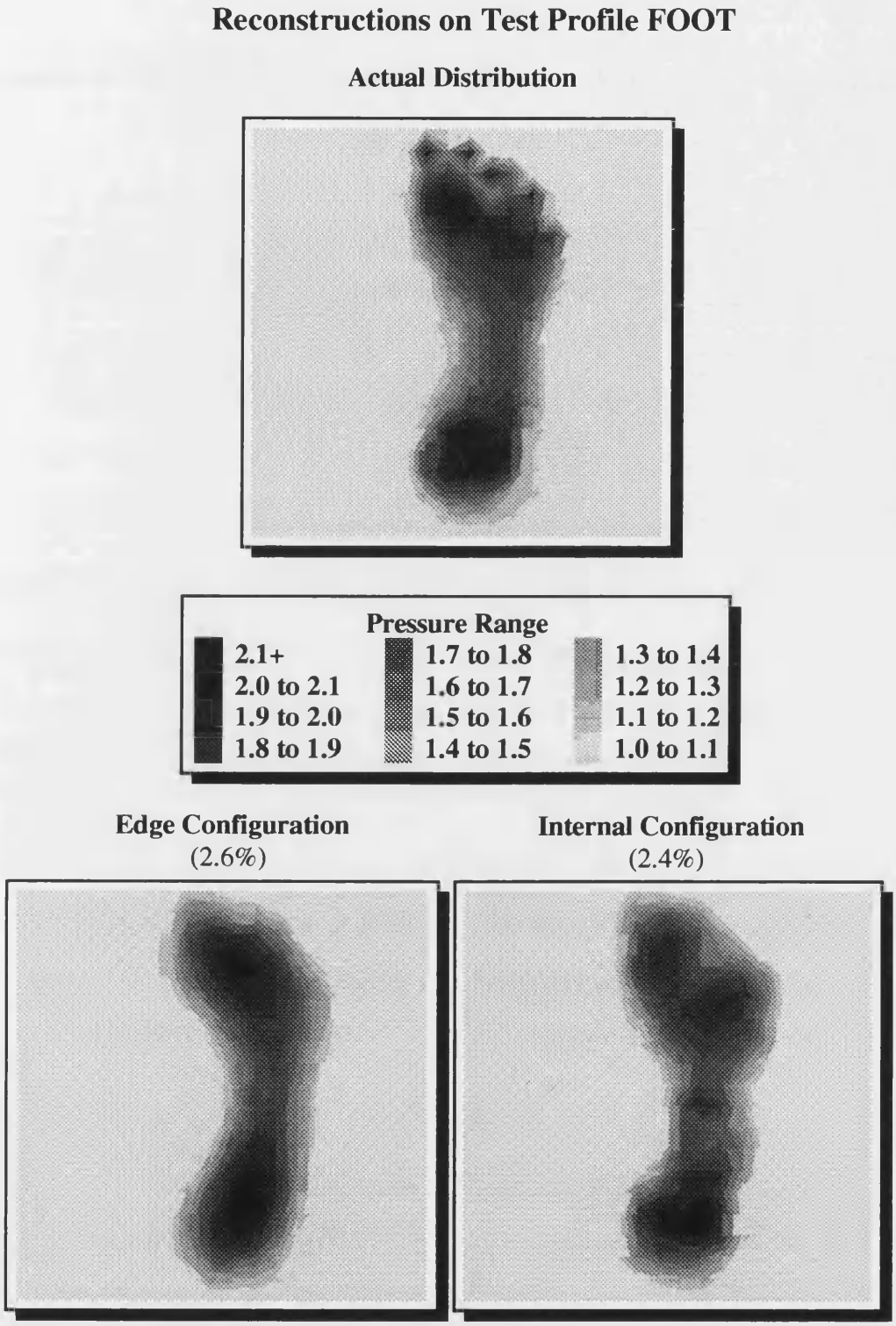


Figure 7.31 Reconstructions on the 400 element mesh using test distribution FOOT comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

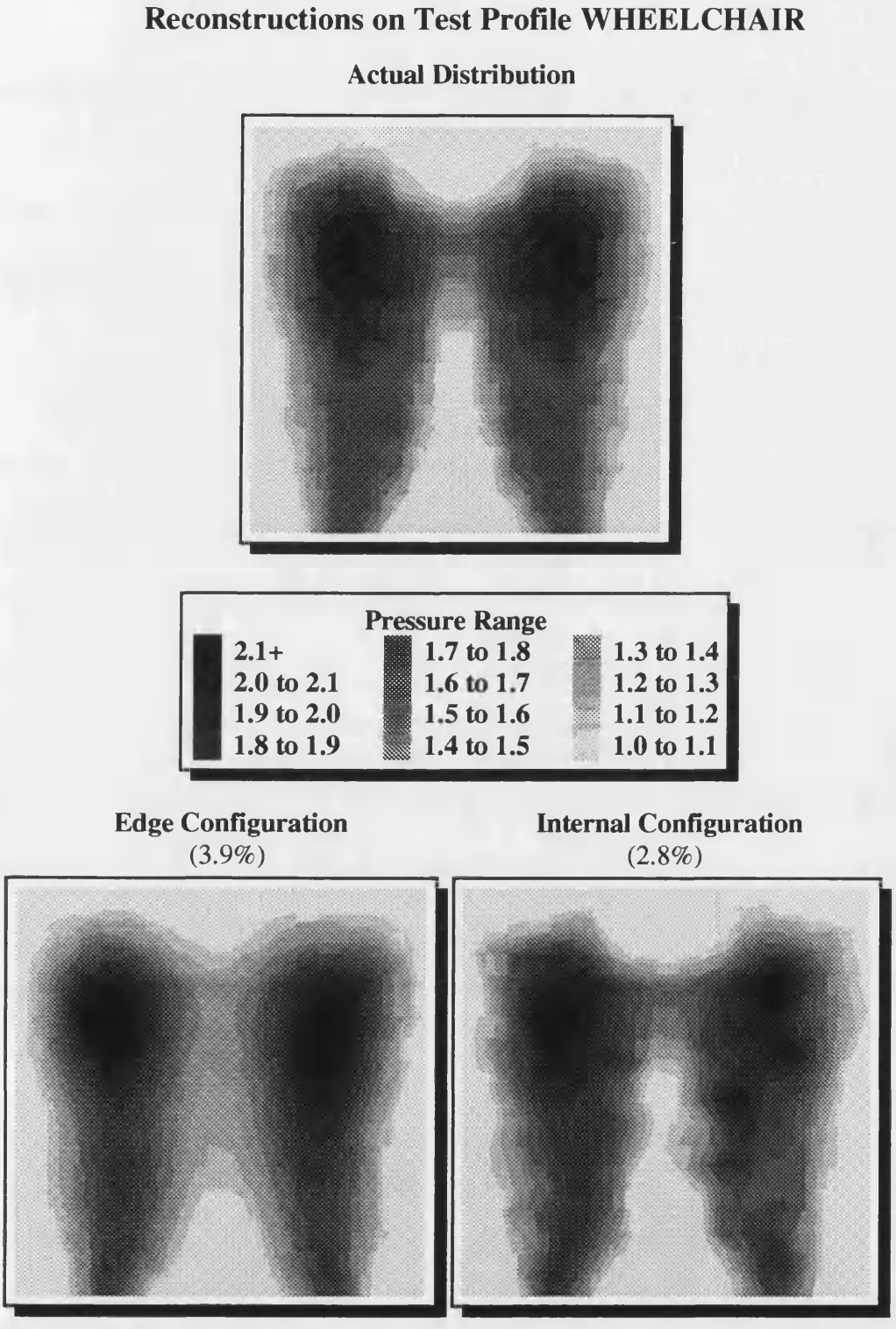


Figure 7.32 Reconstructions on the 400 element mesh using test distribution WHEELCHAIR comparing the Edge and Internal electrode configurations. The element error is shown in brackets.

7.2.3.4. Analysis

A spectral analysis gives an insight into why the Internal configuration is superior to the Edge configuration. The spectrum for all 3 meshes is shown in Figures 7.33 to 7.35. The Internal configuration has a more even spread of Jacobian singular values and the all important higher singular numbers have larger singular values. The improved conditioning results in less error in the resistivity distribution when using the regularised Gauss-Newton method.

The improvements in reconstruction of the Internal over the Edge electrode configuration spans a wide range of voltage measurement errors. The degree of improvement was found to be larger on the smaller meshes as typified by the two graphs in Figures 7.36 and 7.37. This can be attributed to the larger relative differences in conditioning when comparing the 100 element and 400 element electrode placement configurations, as shown in Figures 7.33 and 7.35.

The Concentrated electrode configuration of Section 7.2.3.2 has also been plotted in Figure 7.34 and illustrates how ill conditioned this electrode configuration is. It is, in fact, almost singular to machine precision ($\sim 10^{15}$). For any given electrode configuration, it has been observed that there is very little change in the spectrum for a wide variety of resistivity distributions. There can, however, be a large difference in distinguishabilities, as demonstrated in the Concentrated electrode configuration study. The reconstruction success of a resistivity distribution does not only depend on the conditioning, but also the distinguishability. Summarising, the best electrode configuration will be one that is best conditioned as it will give superior results over a wide range of differing distributions. However, it is not necessarily the optimum configuration for any one particular electrode pattern. This is clear from the Concentrated electrode configuration, which is terribly conditioned, but the TOPLEFT distribution is reconstructed quite accurately.

The distinguishabilities for the pressure distributions of Figures 7.19 to 7.32 are graphed in Figures 7.38 to 7.40. Comparing distinguishabilities between the Edge and Internal configurations, there is not much difference, apart from a few exceptions. The BLOCK profiles have more of a central load, whereas close inspection of the WHEELCHAIR profiles shows that much of the distribution is close to the mat's edge. The

distinguishabilities for the WHEELCHAIR tests are thus better with the Edge electrode configuration, whereas the BLOCK tests have better distinguishabilities with the Internal electrode configuration. Although WHEELCHAIR distinguishabilities are better on the Edge electrode configuration, the problem is not as well conditioned. The better conditioning dominates the distinguishability, resulting in superior results with the Internal electrode configuration. Test profiles FOOT and BUTTOCKS have perturbations both near the edge and towards the centre, resulting in distinguishabilities that are similar on both electrode configurations. Although the distinguishabilities are similar, the superior reconstructions are with the Internal electrode configuration as the conditioning is superior.

The best electrode configuration is one that combines a superior conditioning and superior distinguishability. The TINYCENTRE test distribution of Figure 7.27, is an example of this. It provides one of the best improvements of the Internal configuration over the Edge configuration. It is this type of resistivity distribution that previous EIT systems have not been able to reconstruct accurately. The TINYEDGE test distribution of Figure 7.28, is a repositioned TINYCENTRE distribution and was devised to monitor performance of the Internal configuration with small objects near the edge of the mat. The electrode positioning in relation to this distribution has, in fact, resulted in slightly better distinguishability for the Internal distribution. Consequently, the Edge configuration produces an inferior reconstruction.

One final set of spectrums to examine are those in Figure 7.41. This example shows that the conditioning of the problem is better with 36 electrodes, as opposed to 32. The tests in this thesis always make comparisons using the same number of electrodes. This example also shows that when the number of electrodes is constant (32), the optimal currents provide the best conditioned system with a fixed electrode configuration. The Neighbouring[86] currents shown are one of the first current patterns used by EIT researchers and, by comparison, are poorly conditioned.

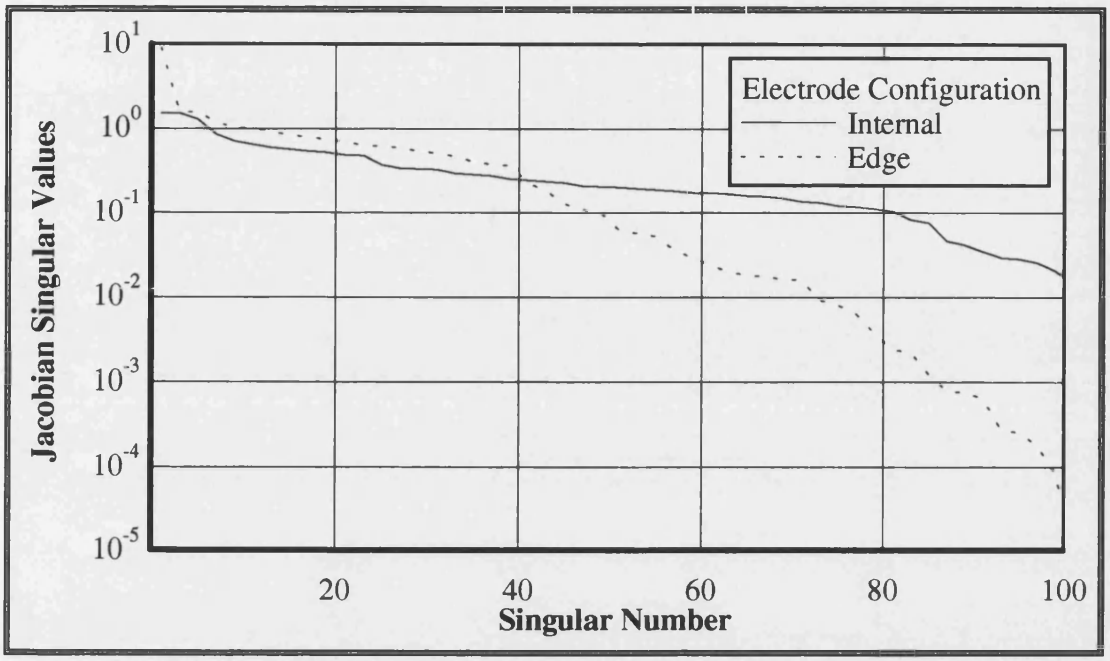


Figure 7.33 Jacobian singular value comparisons on the 100 element mesh

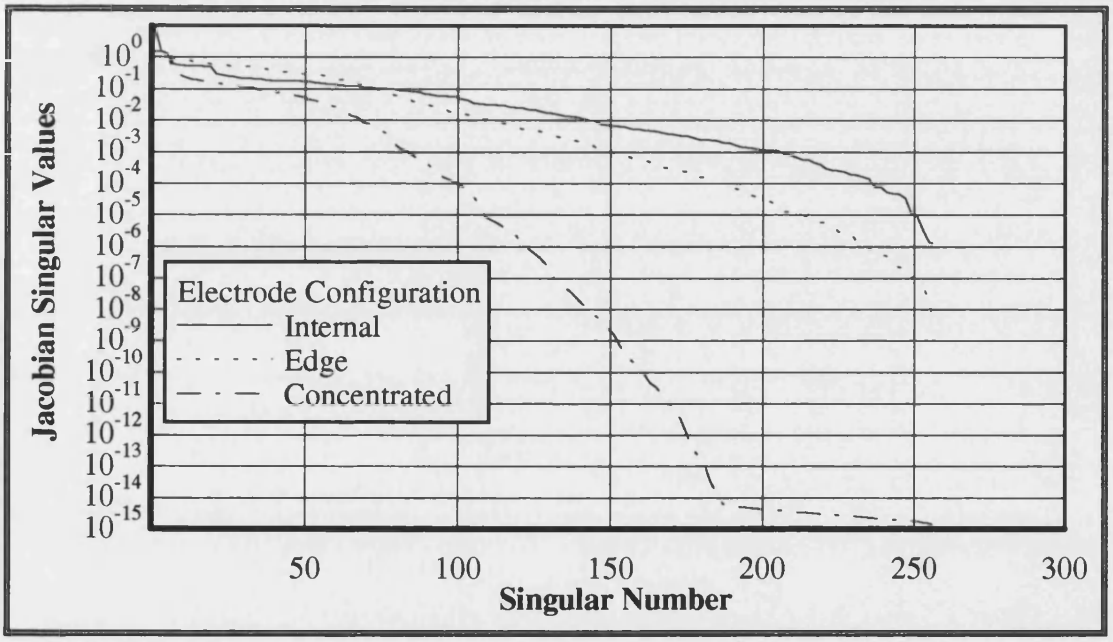


Figure 7.34 Jacobian singular value comparisons on the 256 element mesh using different configurations of 37 electrodes

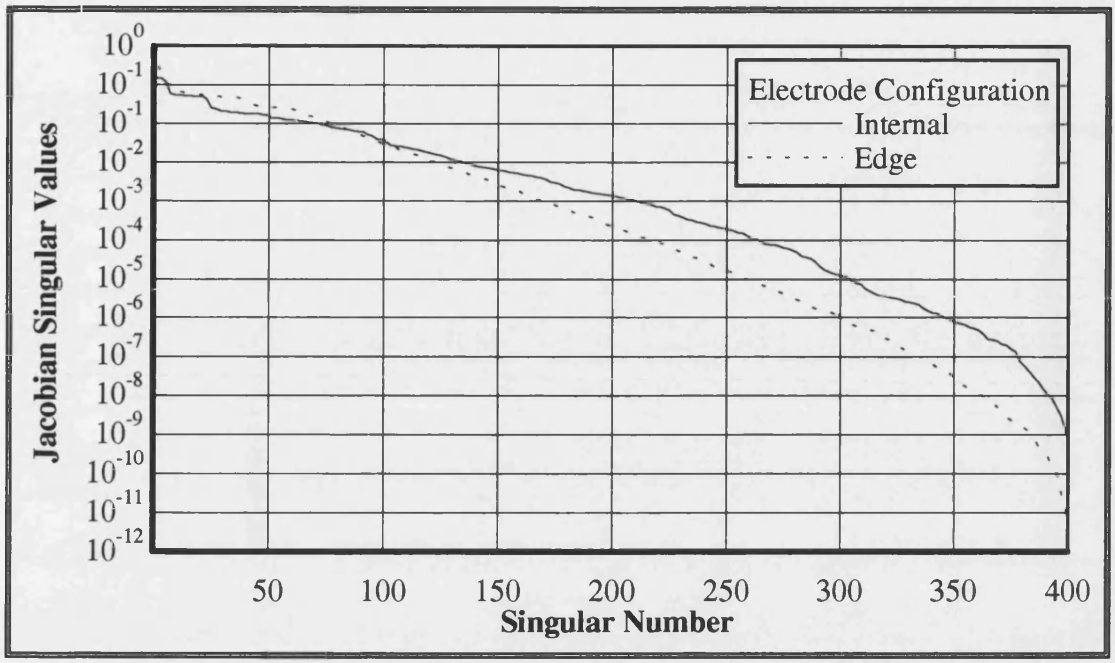


Figure 7.35 Jacobian singular value comparisons on the 400 element mesh

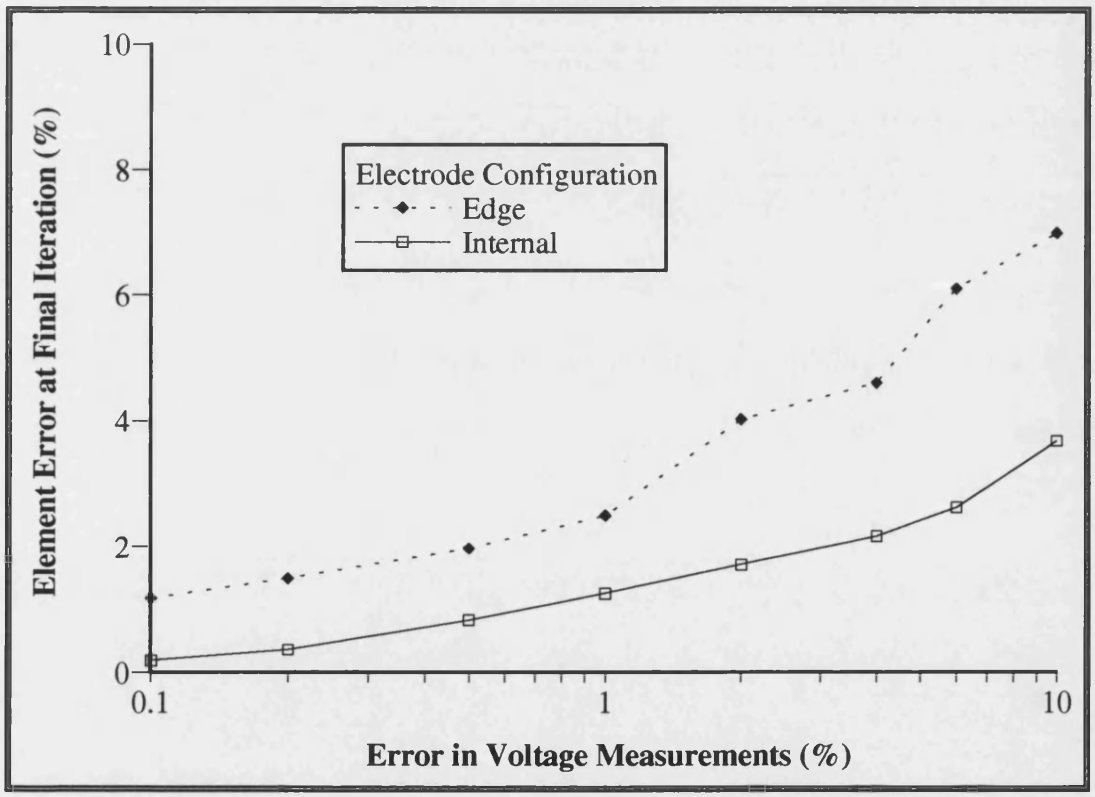


Figure 7.36 Graph of the average error in each element at the final iteration for a range of added voltage measurement errors. Results are for test profile BUTTOCKS on the 100 element mesh.

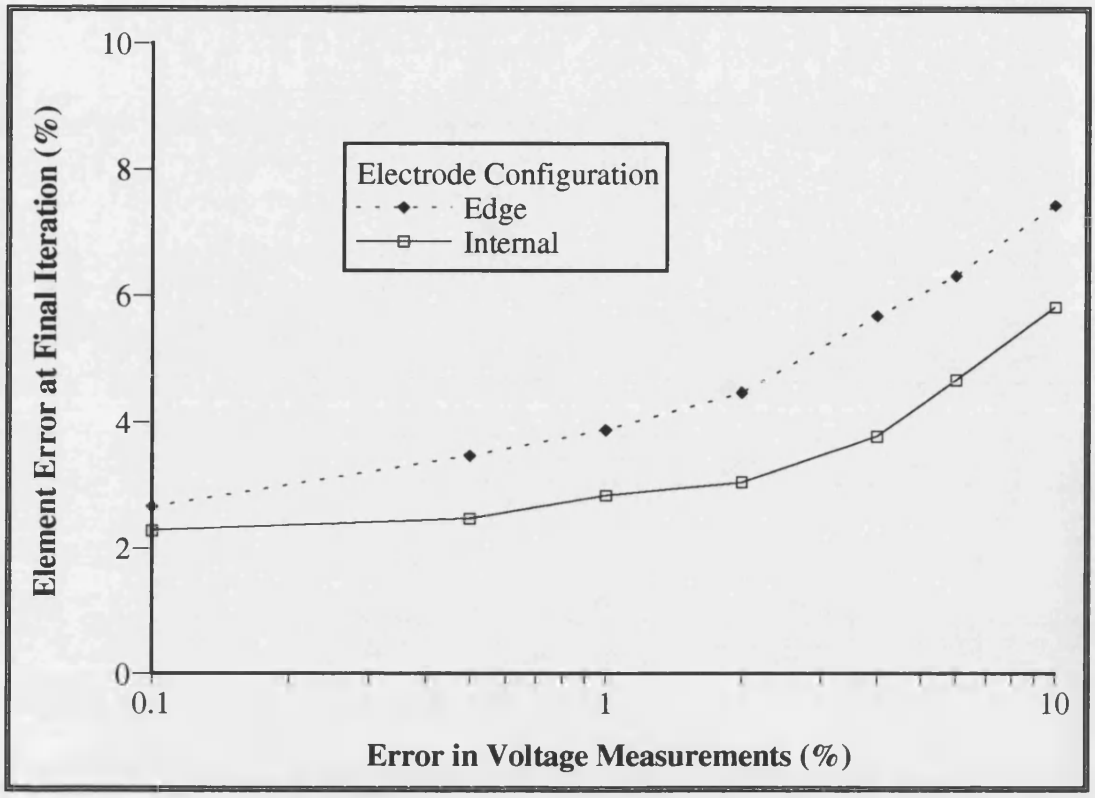


Figure 7.37 Graph of the average error in each element at the final iteration for a range of added voltage measurement errors. Results are for test profile WHEELCHAIR on the 400 element mesh.

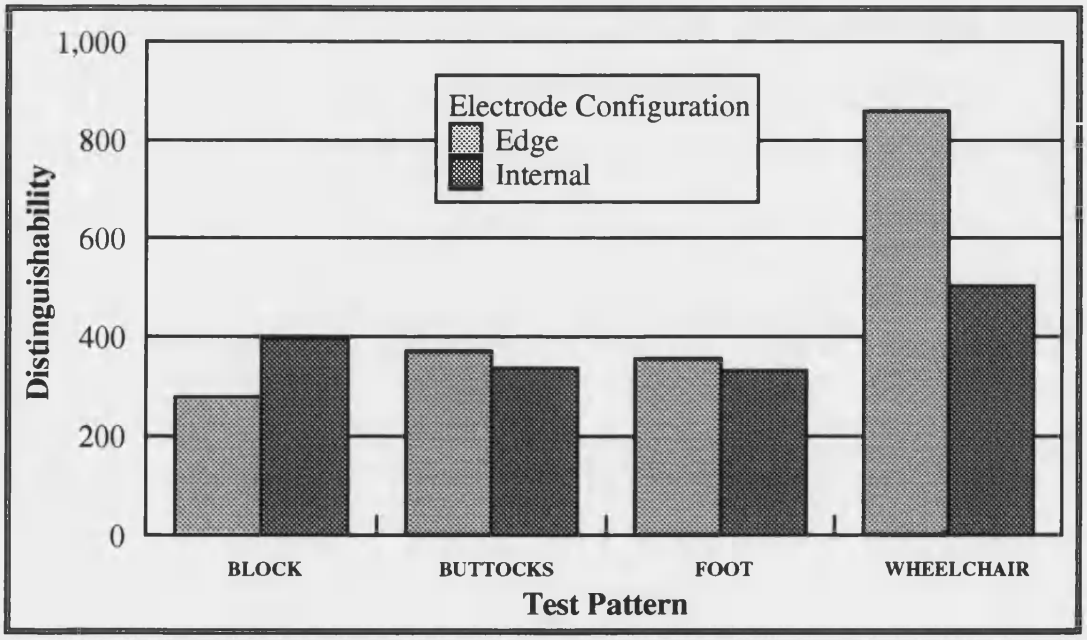


Figure 7.38 Distinguishability comparisons for the 100 element mesh test patterns

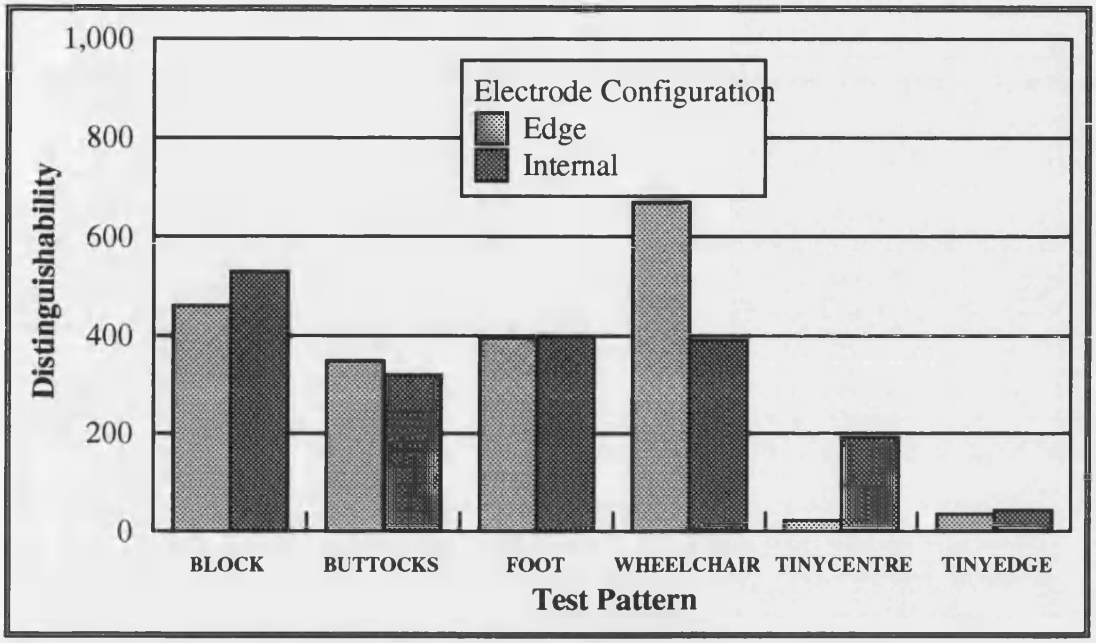


Figure 7.39 Distinguishability comparisons for the 256 element mesh test patterns

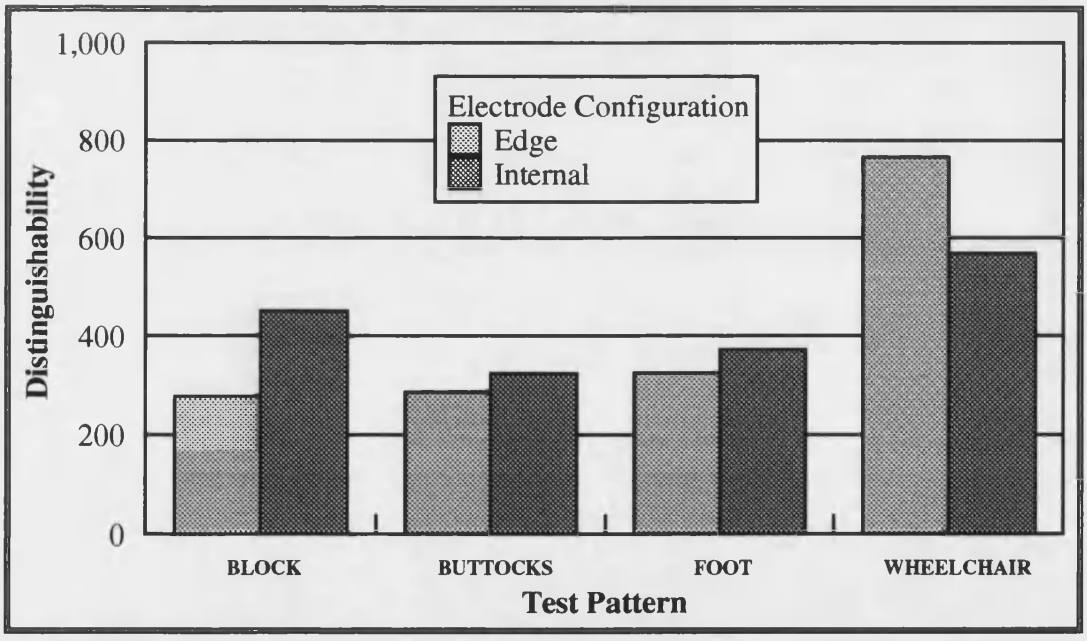


Figure 7.40 Distinguishability comparisons for the 400 element mesh test patterns

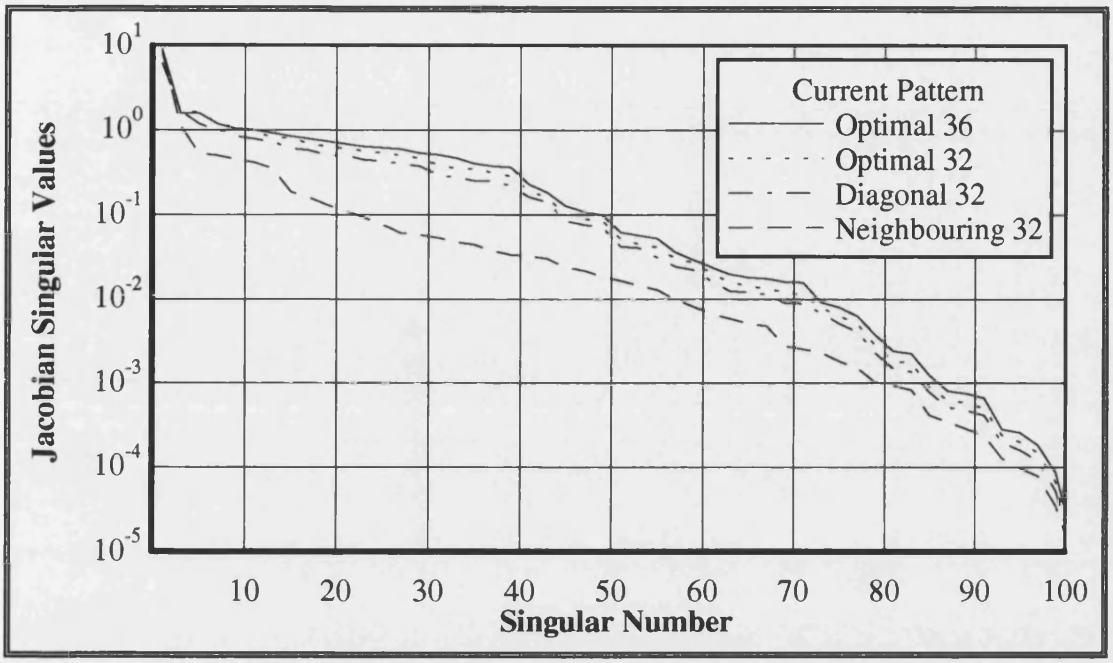


Figure 7.41 Jacobian singular values on the 100 element mesh Edge electrode configuration comparing different current patterns. The number of electrodes used for data collection is shown with the type of current pattern.

7.2.4. Discussion and Conclusions

A change in resistivity in a particular place on the mat will lead to the smallest voltage changes occurring at electrodes that are furthest away. Should some area of the mat have a low density of electrodes, then the system is less sensitive to resistivity changes in this area. Optimal currents on an edge electrode configuration ensures, to a certain extent, that currents flow throughout the domain. Placing electrodes throughout the domain takes this one step further. The intuitive reason for the superiority of internal electrode configurations is that the currents are forced into all parts of the mat. Information can also be measured across the entire mat area. When electrodes are used solely on the edge of the mat, any resistivity changes in the centre will not produce very large voltage measurement changes at the electrodes. The signal to noise ratio in this case is very low and voltage measurement errors swamp the signal. These errors result in the Gauss-Newton method taking a step in the wrong direction. This is kept under control, to some degree, by regularisation, however, too much regularisation smoothes the reconstructed distribution quite considerably. The preferred option is to have the minimum regularisation possible and the use of internal electrodes facilitates this.

A spectral analysis of the Internal and Edge electrode configurations gives an insight into the resulting reconstruction differences. Combining this analysis with an examination of the distinguishability, shows that the Internal configuration has the edge. Not only is this over a wide range of practical pressure profiles, but also over a wide range of measurement errors. It has been shown that the best electrode configuration is one where the electrodes are placed directly beneath the pressure distribution. However, this is an impractical solution for a system that is meant to measure a variety of pressure profiles. The solution has proved to be one with electrodes scattered throughout the mat. The resulting Jacobian matrix conditioning is superior and the distinguishability is good for all pressure distributions, especially small centrally located distributions. The net effect is a more accurate reconstruction.

The inability to distinguish small resistivity distribution in the centre with an edge configuration is a consequence of EIT's low spatial resolution in this region. The centrally located electrodes increases EIT's usually very low spatial resolution in the central region. Consequently, the FEM mesh can be made finer in the important central area, and the smoothing of many pressure distributions no longer occurs.

8. Conclusions and the Future of EIT Interface Pressure Measurement

This chapter is a summary and conclusion of the results presented in this thesis. The clinical usefulness and viability of using an EIT based pressure scanner is analysed. Further work to obtain a working interface pressure scanner is yet to be done and this is briefly outlined.

8.1. Summary and Conclusions

The proposed EIT pressure scanner requires a mat exhibiting known pressure-resistivity characteristics. The preferred solution for this requirement is a carbon fibre felt encapsulated in a gel or silicon rubber. Carbon fibre felt exhibits good electrical characteristics in that it has low hysteresis, is thermally stable and gives repeatable results. The mat would be both thin and flexible, and would conform to the body-support interface. Point electrodes can be accurately attached to any required position on the mat. This solution would also be inexpensive and extremely durable due to the physical properties of carbon fibres.

The pressure sensitive mat is a square domain and is modelled by square bilinear elements. These elements provide the optimal combination of spatial resolution, accuracy and solution time for the EIT system. Cutting the mat to the shape of the domain can be accurately achieved, minimising geometric modelling errors. These elements lend themselves well to sparse matrix methods, as the nested dissection matrix ordering provides a faster solution than the more commonly used minimum degree ordering. The nested dissection ordering is better optimised by the sparse vector method and full exploitation of the lower path lengths has minimised the finite element equations and master stiffness matrix solution time. An accurate Jacobian matrix calculation that is very efficient has been described. This calculation takes full advantage of sparse matrix techniques and removes all redundant evaluations in the master stiffness inverse matrix. Projections have also been chosen, whereby the current source can be floated; this has important repercussions as the grounded node is kept in the same position, thus avoiding recalculation of the master stiffness matrix for each projection. Before these optimisations were implemented, the finite element equation and inverse master stiffness matrix calculations completely dominated the reconstruction. They have now been reduced to a minor part of the regularised Gauss-Newton inverse problem solution time.

A reduction in time has enabled the use of a larger finite element mesh to simulate the mat resistance, thereby increasing the overall image accuracy.

The optimisation technique which most successfully reconstructs the pressure distribution has proved to be the Gauss-Newton method regularised by second finite differences. This was shown in the special EIT application of pressure measurement, where the electrode positions and element type can be defined by the designer. The choice of square bilinear elements enables easy implementation of the superior second finite difference regularisation. A good starting estimate can be provided for the regularised Gauss-Newton method as the no-load resistivity is accurately known. These combinations result in good pressure reconstructions, albeit with a slight smoothing effect. The smoothing effect removes details such as the toes on the foot. However, as far as the medical field is concerned, this sort of resolution has never been achieved and is not necessarily needed. The proposed pressure scanner provides enough detail for many useful environments, for example, it obtains the peak pressure distributions of the bony protuberances under the buttocks.

The practical application of EIT to a pressure measuring environment has accuracy advantages over the more usual EIT applications. These are the flowing of currents in two-dimensions and accurate placing of the electrodes, however, there will be errors in the voltage readings. These will be due to slight inaccuracies in electrode placement, finite element modelling errors, noise in the electronics and quantisation error when digitising the voltages. Reconstruction in a real-world environment turns the optimisation problem into a large residual problem. Regularisation is needed when the residuals are not small. The regularisation parameter method that consistently gave good results was the Marquardt choice, however, this choice requires many iterations to solution. Slightly inferior, but faster reconstructions can be obtained when using the dwindling parameter method.

The optimum electrode configuration for practical pressure distributions in the medical field has been studied. The best configuration proved to be one with the electrodes distributed throughout the mat. The resulting superior conditioning is a consequence of the currents being forced into all areas of the mat and the accessibility of the central area for voltage measurement. The net effect is successful reconstruction over a wide range of pressure distributions and over a wide range of voltage measurement errors. Compared to a configuration of electrodes on the edge of the mat, the most impressive

improvements are those with small pressure profiles in the centre of the mat. Regularisation is a form of solution stabilising technique, but does not remove the inherent ill-conditioning. The only effective means of combating ill-conditioning is to improve the system conditioning and this has been achieved through the use of internal electrodes.

8.2. Viability of an EIT Pressure Scanner

This thesis has presented all the theory and solutions for implementing a real-world interface pressure scanner. Computer simulations using practical voltage measurement errors have shown that reconstructions are possible and that they are fairly accurate. More importantly, they are usable by the medical profession for pressure sore analysis and could be incorporated into a pressure monitoring device. A single reconstruction currently takes between half a second and one minute, depending on the resolution used. Constantly improving computing power and further algorithm optimisations, such as those listed in the following section, will ensure that a number of reconstructions will be possible per second. High resolution analysis will then be possible in real time, rather than off-line.

The practical voltage measurement error used throughout has been set to a conservatively large figure of 1%. With accurate electrode placement and low noise electronic design, it is most likely that measurement errors will be lower; the consequence will be superior reconstructions than those shown in this thesis.

No pressure scanner can be implemented without a material that exhibits the proposed pressure-resistivity characteristics. Moreover, a material which has low hysteresis, gives repeatable and accurate results in a practical environment is required. A carbon fibre based mat fulfils these requirements, thereby ensuring the viability of an EIT based pressure scanner.

8.3. Future Work Required for an EIT Pressure Scanner

The viability of using EIT techniques in a practical pressure scanner has been established. The computer simulations must now be implemented into a successful working pressure scanner; everything is in place to achieve this. The reconstruction algorithms work well and the important pressure sensitive material has been found. Carbon fibre felt is inexpensive to manufacture and the next step would be to obtain

samples. Prototype electronics to inject currents and measure the resulting voltages must be built to interface with a PC. Clinical trials would then fully assess a working system.

Large reconstruction speed optimisations have been presented in this thesis. Previously, no results on meshes larger than 25 elements could be obtained as the reconstruction was too lengthy. Presently, the solution of the least squares equation dominates the reconstruction speed. A faster solution of these equations may be possible through a sparse implementation of Hachtel's method. Paulson[50] describes a more promising method which reduces the size of the least squares equation, resulting in a faster reconstruction. This method is based on the use of optimal current patterns as well as optimal voltage measurement patterns. Currently, the starting estimate for a reconstruction is the no-load resistivity. Fewer iterations will be possible when the previous reconstruction is used as the initial estimate.

Seagar[51] analysed the maximum or best theoretical spatial resolution obtainable with edge electrodes only. The best resolution turned out to be zero at the edge and just 50% at the centre of the domain. The resolution at the centre is incredibly high and ultimately necessitates the use of huge FEM elements in the centre of the mat. Theoretical limits could also be studied for an internal electrode configuration. This would define a limit for mesh refinement between electrodes, in addition to the limit defined by equation (4.2).

Should a silicon glue be used to attach the electrodes to the mat, the contact impedance of these point electrodes will be low. However, there are other electrode contact effects which have been studied when modelling the contact impedance. Hua[49] developed a finite element model for the contact impedance as he used large electrodes. Large electrodes used by Gisser[90] achieved a more uniform distribution of current through the domain, i.e. they improved the distinguishability. A large electrode differs from a point electrode in that it has shunting and edge effects which cannot be modelled by a single resistor. The FEM models developed, took account of these effects. A study on large electrodes will conclude whether or not they will be necessary for use in an internal electrode configuration.

8.4. Pressure Measurement in the Clinical Environment

Pressure sores are for many, an unfortunate consequence of being hospitalised or disabled in some form. They generally occur in patients who are in operating theatres for

the short operating time needed, in the elderly, the physically injured and disabled. Most of these pressure sores need not occur if the correct care is given to the patient. However, with the increasing demands made on nursing staff, the recommended patient turning of at least every two hours, is not strictly adhered to. The consequent pressure sore formation is a huge drain on hospital finances. The prevention of pressure sores is of utmost importance in the management of patient care and it has been established that any device to aid this will be most welcomed by the medical field and the sufferers of such sores.

If a good body-support interface pressure monitoring device is cheap enough, it would be incorporated into most hospital wards throughout the country, to monitor patients who are most at risk of sore formation. Pressure measurement data is also needed for research purposes, to aid wheelchair cushion designers and to evaluate specialised beds and wheelchairs. Other medical uses for interface pressure measurement include gait analysis, denture alignment, orthopaedic footwear design, posture analysis and in the control of paraplegic walking. The interface pressure measurement devices that are available, do not fulfil medical requirements, mainly in that they are inaccurate, interfere with the body-support surface interface pressure characteristics and only measure pressure at certain points. The requirement is for a thin, flexible mat to map distributed pressure over a region. The proposed EIT measurement system fulfils this criterion and the viability of measuring practical pressure distributions has been shown. It will be capable of monitoring pressure over time to obtain the important pressure-time characteristics needed in aiding pressure sore prevention. Although pressure measurement does not completely characterise the body-support interface, its clinical usefulness and necessity in assessing potential tissue damage by any support surface, is most clearly established.

Acknowledgements

I would like to thank Dr Dick Lipczynski for supervision and guidance throughout the project. Thanks are also due to Professor JF Eastham and the School of Electronic and Electrical Engineering for provision of facilities. The project is funded by the Science and Engineering Research Council whose kind provision of finances has made the research possible.

Thanks must also be extended to Lindsay Grant of the Medical Physics Department at the Royal United Hospital, Bath and Rob Knight for valuable advice and assistance. Completing the research would not have been an enjoyable experience without the company of my fellow postgraduate colleagues, Alan Gregory, Sean Hagan, Salem Mebarki, Andrew Moran, Adrian Ward and John Williams. I am grateful to Chris Murphy for writing the postscript routines for displaying the sparse matrices, and for his and Joy Durkin's proof reading of the thesis. My family and friends, too numerous to name, have all been wonderfully supportive and encouraging over the last few years. Many heartfelt thanks to you all.

References

[1] Knight RA 1991 *The use of EIT techniques to measure interface pressure* PhD Thesis, University of Bath, Bath, BA2 7AY, UK

[2] Holstein P, Nielsen PE and Barras JP 1979 Blood flow cessation at external pressure in skin of normal human limbs: photoelectric recordings compared to local intraarterial blood pressure *Microvasc Res* **17** 71-79

[3] Exton-Smith N 1987 The patients not for turning *Nursing Times* **83** (October 21) 42-44

[4] Garber SL 1986 The role of technology in tissue management *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **8** 1835-1836

[5] McClemont EJW 1984 Pressure Sores *A supplement to Nursing Times* **21**

[6] Versluyen M 1986 How elderly patients with femoral fracture develop pressure sores in hospital *British Medical Journal* **292** 1311-1313

[7] Grant LJ 1985 Interface pressure measurement between a patient and support surface *Care - The British Journal of Rehabilitation and Tissue Viability* **1** 7-9

[8] Specter TH 1990 A Tactile Sensing System for Robotic Manipulation *International Journal of Robotics Research* **9** 25-36

[9] Webster JG 1989 A Pressure Mat for Preventing Pressure Sores *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **11** 1485-1486

[10] Barbenel JC and Sockalingham S 1990 Devices for measuring soft tissue interface pressures *J. Biomed. Eng.* **12** 519-522

[11] Babb FC, Bourland JD, Graber GP, Jones JT and Schoenlein WE 1990 A pressure -sensitive mat for measuring contact pressure distributions of patients lying on hospital beds *Biomed. Instrum. Technol (USA)* **24** 363-370

[12] Reddy NP, Palmieri V, Van G and Cochran B 1984 Evaluation of transducer performance for buttock-cushion interface pressure measurements *Journal of Rehabilitation Research and Development* **21** 43-50

[13] Lindan O, Greenway RM and Piazza JM 1965 Pressure distribution on the surface of the human body: Evaluation in Lying and sitting positions using a 'Bed of springs and Nails' *Archives of Physical Medicine and Rehabilitation* **46** 378-385

[14] Barbenel JC 1983 *Measurement of interface pressures: Pressure Sores pp 67-78* (Ed. Barbenel JC, Forbes CD and Lowe GDO, London, MacMillan, ISBN 0-333 31889-7)

[15] Sakamoto R 1987 Electrically-conductive rubbers with carbon black. 3. Electrically conductive rubber. *International Polymer Science and Technology* **14** 40-47

[16] Nagata M 1985 Metal-filled conductive rubbers. 1. Pressure-sensitive rubbers. *International Polymer Science and Technology* **13** 44-49

[17] Mokshagundam A 1988 *Conductive Elastomers and Carbon Fibers. In: Tactile Sensors for Robotics and Medicine.* (Ed: JG Webster, John Wiley & Sons, New York) 125-148

[18] Kost J, Narkis M and Foux A 1984 Resistivity Behavior of Carbon-Black-Filled Silicone Rubber in Cyclic Loading Experiments *J.Appl.Polym.Sci.* **29** 3937-3946

[19] Basarab-Horwath I and Faraj M 1993 *A Prototype Distributed Pressure Sensor Utilising Electrical Impedance Tomography: Initial Results. In: Tomographic Techniques for Process Design and Operation* (Eds: Beck MS, Campogrande E, Morris M, Williams RA, Waterfall RC, Computational Mechanics Publications, Southampton)

[20] Katahira K 1979 A new type of force-sensitive device using carbon fiber and its biomedical application *Fukushima J.Med.Sci.* **26** 121-132

[21] Pruski A and Mutel B 1983 Capteur sensoriel pour robot industriel. *Congres ACFET Productique et Robotique Intelligente, Bescanon, France* 287-295

[22] Larcombe MHE1981 Carbon Fibre Tactile Sensors *Proc.1st.Int.Conf.Robot Vision Sensory Control* 273-276

[23] Katahira K, Tsukahara S, Fujiwara M and Miura H 1984 Carbon-Fiber Foot-Force Sensor System for Gait Analysis *Fukushima J.Med.Sci.* **30** 29-37

[24] Robertson BE and Walkden AJ 1986 *Tactile sensor system for robotics. In: Robot Sensors Volume 2 Tactile & Non-Invasive* (Ed: A Pugh, IFS (Publications) Ltd, UK)

[25] Pruski A and Mutel B 1985 Direct Contact Sensors Based on Carbon Fibre *Proc.Int.Conf.Robotics and Factories of the Future* (Ed. SN Dwivedi, Springer-Verlag, Berlin) 409-415

[26] Beck, M.S., Campogrande, E.C., Morris, M., Williams, R.A., Waterfall, R.C. 1993 *Tomographic Techniques for Process Design and Operation* (Computational Mechanics Publications, Southampton)

[27] Barber DC 1989 A review of image reconstruction techniques for electrical impedance tomography *Med. Phys.* **16** 162-169

[28] Bates RHT, Garden KL and Peter TM 1983 Overview of computerized tomography with emphasis on future developments *Proc IEEE* **71** 356-372

[29] Henderson RP and Webster RG 1978 An impedance camera for spatially specific measurements of the thorax *IEEE Trans. Biomed. Eng.* **BME-25** 250-254

[30] Bates RHT McKinnon GC and Seagar AD 1980 A limitation on systems for imaging electrical conductivity distributions *IEEE Trans. Biomed. Eng.* **BME-27** 418-420

[31] Isaacson D, Cheney M and Newell JC 1992 Comments on reconstruction algorithms *Clin. Phys. Physiol. Meas.* **13 Suppl A** 83-89

[32] Barber DC and Brown BH 1984 Applied Potential Tomography *J. Phys. E: Sci. Instrum.* **17** 723-732

[33] Brown BH, Barber DC and Seagar AD 1985 Applied Potential Tomography: Possible clinical applications *Clin. Phys. Physiol. Meas.* **6(2)** 109-121

[34] Woo EJ, Hua P, Webster JG and Tompkins WJ 1992 Measuring lung resistivity using Electrical Impedance Tomography *IEEE Trans. Biomed. Eng.* **BME-39** 756-760

[35] Harris ND, Suggett AJ, Barber DC and Brown BH 1987 Applications of Applied Potential Tomography (APT) in respiratory medicine *Clin. Phys. Physiol. Meas.* **8 Suppl A** 155-165

[36] Newell JC, Isaacson D, Saulnier GJ, Cheney M and Gisser DG 1993 Acute Pulmonary Edema Assessed by Electrical Impedance Imaging *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **15** 92-93

[37] Yamashita Y and Takahashi T 1981 Method and feasibility of estimating impedance distribution in the human torso *Proceedings of the Vth International Conference on Electrical Bioimpedance (ICEBI)* 87-90

[38] Kim Y, Webster JG and Tompkins WJ 1983 Electrical impedance imaging of the thorax *Journal of Microwave Power* **18** 245-257

[39] Yorkey TJ, Webster JG and Tompkins WJ 1987 An improved pertubation technique for Electrical Impedance Imaging with some criticisms *IEEE Trans. Biomed. Eng.* **BME-34** 898-901

[40] Yorkey TJ, Webster JG and Tompkins WJ 1987 Comparing reconstruction algorithms for Electrical Impedance Tomography *IEEE Trans. Biomed. Eng.* **BME-34** 843-852

[41] Hu M-Q, Hua P and Webster JG 1987 Measurement error in an Electrical Impedance Tomography system *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **9** 1425-1426

[42] Gencer GN, Ider YZ and Nakiboglu B 1990 An alternative solution for the problem of electrode position determination in Electrical Impedance Tomography *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **12** 128-129

[43] Hua P and Woo EJ 1990 *Reconstruction Algorithms. In: Electrical Impedance Tomography* (Ed: Webster JG. Bristol: Adam Hilger)

[44] Huebner KH and Thornton EA 1975 *The Finite Element Method for Engineers 2nd Edition* (John Wiley & Sons, New York, ISBN 0-471-09159-6)

[45] Mitra SK 1969 *Analysis and synthesis of linear active networks* (New York: John Wiley & Sons)

[46] Yorkey TJ 1986 *Comparing reconstruction methods for Electrical Impedance Tomography* PhD dissertation, Department of Electrical and Computing Engineering, University of Wisconsin, Madison, USA

[47] Yorkey TJ 1990 Electrical Impedance Tomography with piecewise polynomial conductivities *Journal of Computational Physics* **91** 344-360

[48] Strang G and Fix GJ 1973 *An Analysis of the Finite Element Method* (Prentice-Hall, Englewood Cliffs, NJ, ISBN 0-13-032946-0)

[49] Hua P 1990 *Modeling and reconstruction methods for Electrical Impedance Tomography* PhD thesis, Department of Electrical and Computing Engineering, University of Wisconsin, Madison, USA

[50] Paulson KS 1992 *Parallel Algorithms for Three Dimensional Electrical Impedance Tomography* PhD Thesis, School of Computing and Mathematical Sciences, Oxford Brookes University, Oxford

[51] Seagar AD, Barber DC and Brown BH 1987 Theoretical limits to sensitivity and resolution in impedance imaging *Clin. Phys. Physiol. Meas.* **8 Suppl A** 13-31

[52] Burnett DS 1987 *Finite Element Analysis* (Addison-Wesley, Reading, Massachusetts, ISBN 0-201-10806-2)

[53] Duff IS, Erisman AM, Reid JK 1986 *Direct Methods for Sparse Matrices* (Oxford University Press, New York, ISBN 0-19-853408-6)

[54] George A and Liu JWH 1981 *Computer Solutions of Large Sparse Positive Definite Systems* (Prentice-Hall, New Jersey)

[55] Numerical Algorithms Group 1991 *NAG Fortran library manual - Mark 15* (Oxford Publishers)

[56] Lewis JG 1982 Implementation of the Gibbs-Poole-Stockmeyer and Gibbs-King Algorithms *ACM Trans. Math. Softw.* **8** 180-194

[57] Tinney WF and Walker JW 1967 Direct Solutions of Sparse Network Equations by Optimally Ordered Triangular Factorization *Proc. IEEE* **55** 1801-1809

[58] George A and Liu JWH 1989 The Evolution of the Minimum Degree Ordering Algorithm *SIAM Review* **31** 1-19

[59] George A 1973 Nested Dissection of a Regular Finite Element Mesh *SIAM J.Numer.Anal.* **10** 345-363

[60] Duff IS, Erisman AM, Reid JK 1976 On George's Nested Dissection Method *SIAM J. Numer. Anal.* **13** 686-695

[61] George A and Liu JWH 1978 An Automatic Nested Dissection Algorithm for Irregular Finite Element Problems *SIAM J.Numer.Anal.* **15** 1053-1069

[62] Tinney WF, Brandwajn V and Chan SM 1985 Sparse vector methods *IEEE Transactions on Power Apparatus and Systems* **PAS-104** 295-301

[63] Gomez A and Franquelo LG 1988 Node ordering for sparse vector method improvement *IEEE Transactions on Power Systems* **PWRS-3** 73-79

[64] George A 1977 Numerical Experiments Using Dissection Methods to Solve n by n Grid Problems *SIAM J.Numer.Anal.* **14** 161-179

[65] Woo EJ 1990 *Computational Complexity. In: Electrical Impedance Tomography* (Ed: Webster JG. Bristol: Adam Hilger)

[66] Adby PR and Dempster MAH 1974 *Introduction to Optimisation Methods* (Halsted Press, John Wiley & Sons, New York, ISBN 0 412 11040 7)

[67] Swann WH 1972 *Direct Search Methods. In: Numerical Methods for Unconstrained Optimisation* (Ed: W Murray Academic Press, London, ISBN 0-1-512250-0)

[68] Woo EJ 1990 *Finite Element Method and Reconstruction Algorithms in Electrical Impedance Tomography* PhD Thesis, University of Wisconsin, Madison, USA

[69] Gill PE, Murray W and Wright MH 1981 *Practical Optimization* (Academic Press, London. ISBN 0-12-283950-1)

[70] Brewer JW 1981 Kronecker products and matrix calculus in system theory *IEEE Transactions on Circuits and Systems* **CAS-25** 772-781

[71] Jennings A and McKeown 1992 *Matrix Computation Second Edition* (John Wiley & Sons, Chichester, ISBN 0 471 93598 0)

[72] Gill PE and Murray W 1978 Algorithms for the Solution of the Nonlinear Least-Squares Problem *SIAM J.Numer.Anal.* **15** 977-992

[73] Artola J and Dell J 1994 Broyden quasi-Newton method applied to electrical impedance tomography *Electron.Lett.* **30** 27-28

[74] Marquardt DW 1963 An Algorithm for Least-Squares Estimation of Nonlinear Parameters *J.Soc.Indust.Appl.Math.* **11** 431-441

[75] Dobson DC 1992 Convergence of a Reconstruction Method for the Inverse Conductivity Problem *SIAM.J.Appl.Math.* **52** 442-458

[76] Moré JJ 1978 *The Levenburg-Marquardt Algorithm: Implementation and Theory. In: Lecture Notes in Mathematics 630* (Ed: GA Watson, Springer Verlag, Berlin) 105-116

[77] Morozov VA 1984 *Methods for Solving Incorrectly Posed Problems* (Springer Verlag, Berlin, ISBN 3-540-96059-7)

[78] Hua P, Webster JG and Tompkins WJ 1988 A regularised Electrical Impedance Tomography algorithm *Clin. Phys. Physiol. Meas.* **9 Suppl A** 137-141

[79] Breckon WR and Pidcock MK 1988 Data errors and reconstruction algorithms in electrical impedance tomography *Clin. Phys. Physiol. Meas.* **9 Suppl A** 105-109

[80] Levenburg K 1944 A Method for the Solution of Certain Non-Linear Problems in Least Squares *Quart.J.Appl.Math.* **2** 164-168

[81] Davies M and Whitting IJ 1971 *A Modified Form of Levenburg's Correction. In: Numerical Methods for Non-linear Optimisation* (Ed: FA Lootsma, Academic Press, London) 191-201

[82] Dennis JE and Schnabel RB 1983 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, London, ISBN 0-13-627216-9)

[83] Woo EJ, Webster JG and Tompkins WJ 1990 The improved Newton-Raphson method and its parallel implementation for static impedance imaging *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **12** 102-103

[84] Duff IS and Reid JK 1976 A Comparison of Some Methods for the Solution of Sparse Overdetermined Systems of Linear Equations *J.Inst.Maths.Applics* **17** 267-280

[85] Moré JJ, Garbow BS and Hillstrom KE 1980 User guide for MINPACK-1 *Argonne National Labs Report ANL-80-74*

[86] Brown BH and Seagar AD 1987 The Sheffield data collection system *Clin. Phys. Physiol. Meas.* **8A** 91-97

[87] Hua P, Webster JG and Tompkins WJ 1987 Effect of the Measurement Method on Noise Handling and Image Quality of EIT Imaging *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol Soc.* **9** 1429-1430

[88] Isaacson D 1986 Distinguishability of conductivities by electric current computed tomography *IEEE Trans. Medical Imaging* **MI-5** 91-95

[89] Golub GH and Van Loan CF 1989 *Matrix Computations* (Johns Hopkins University Press, London, ISBN 0-8018-3772-3)

[90] Gisser DG, Isaacson D and Newell JC 1987 Current topics in impedance imaging *Clin. Phys. Physiol. Meas.* **8 Suppl A** 39-46

[91] Hua P, Woo EJ, Webster JG and Tompkins WJ 1992 Improved Methods to Determine Optimal Currents in Electrical Impedance Tomography *IEEE Trans. Med. Imaging.* **MI-11** 488-495

[92] Fulton WS and Lipczynski RT 1993 Body-Support Pressure Measurement Using Electrical Impedance Tomography *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **15** 98-99

[93] Moskowitz MJ, Paulsen KD and Ryan TP 1993 Thermal images obtained by a combined invasive-noninvasive approach using Electrical Impedance Tomography *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **15** 88-89

[94] Pilkington T, Eyüboglu M and Wolf P 1989 Utilization of Esophageal Reference Electrode to Enhance Impedance Imaging *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **11** 482

[95] Stuart GW 1973 *Introduction to Matrix Computations* (Academic Press, London)

[96] Silvester PP and Ferrari RL 1990 *Finite elements for electrical engineers. 2nd Edition.* (Cambridge University Press, Cambridge)

[97] Tong P and Rossettos JN 1977 *Finite Element Method: Basic Technique and Implementation* (MIT Press, London, ISBN 0-262-20032-5)

Appendix A

A1. Graph Theory

A few basic graph theory concepts aid in visualising matrix manipulations and describing matrix algorithms. Some of the basics required for this thesis are presented here, but further analysis for sparse matrices may be found in George[54] and Duff[53].

A1.1. Basic Graph Theory and Terminology

A symmetric matrix can be represented by an **undirected graph**, also known as a **graph**. A graph consists of a set of **nodes** and **edges** between nodes. An example is pictured in Figure A1.1(a). For each non-zero entry in the matrix there is an associated edge in the graph. For a non-zero entry a_{ij} in any symmetric matrix **A**, there is an implied non-zero entry a_{ji} and an edge from node i to node j. The edge is represented diagrammatically as a line between the relevant nodes. For example, the entry a_{24}, in Figure A1.1(a) has an edge between nodes 2 and 4.

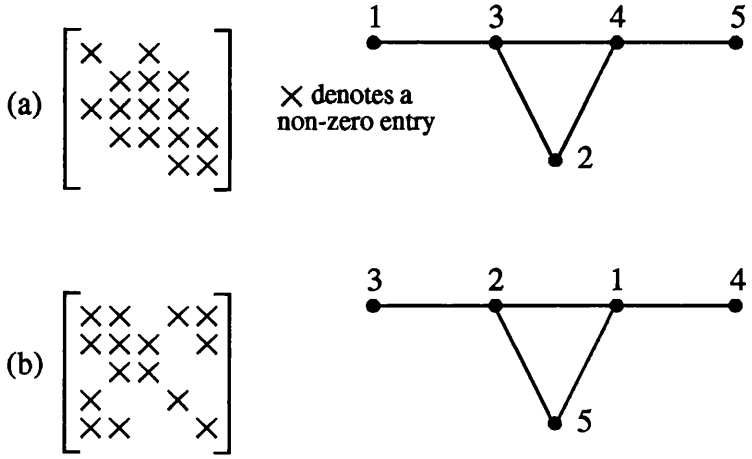


Figure A1.1(a) A symmetric matrix example and its associated graph

(b) Re-ordered matrix and its associated graph

The aim in using sparse matrix theory is to find a matrix ordering to reduce the matrix storage and solution times. In graph theory, a re-ordered matrix has the same structure as the unordered matrix, but the node numbering is different. This is illustrated in

Figure A1.1(b), a re-ordering of the matrix shown in Figure A1.1(a). So the aim of finding a good matrix ordering, corresponds to labelling the associated graph. This is what the minimum degree algorithm, for example, does.

The finite element matrix used in this thesis is a full matrix and its associated graph is shown in Figure A1.2.

Figure A1.2 Finite element matrix and associated graph

In the two examples already given, there are loops in the graphs. A special case can occur when there are no closed loops in the graph. This particular case is when the graph is a rooted tree and is illustrated in Figure A1.3(a).

The **degree** number of a node *i* can be seen as the number of connected edges to *i* or the total number of off-diagonal non-zeros. For example, in Figure A1.3(a), node 1 has degree 3 and node 5 has degree 1.

Graphs are good for showing what happens when adding or subtracting matrix rows. Referring to the matrix in Figure A1.3(a), row 3 can be altered by subtracting row 2 from it and similarly column 3 can be altered by subtracting column 2 from it. This produces the graph in Figure A1.3(b). Notice how a new edge from node 1 to node 3 has been produced, corresponding to the new entries a_{13} and a_{31}.

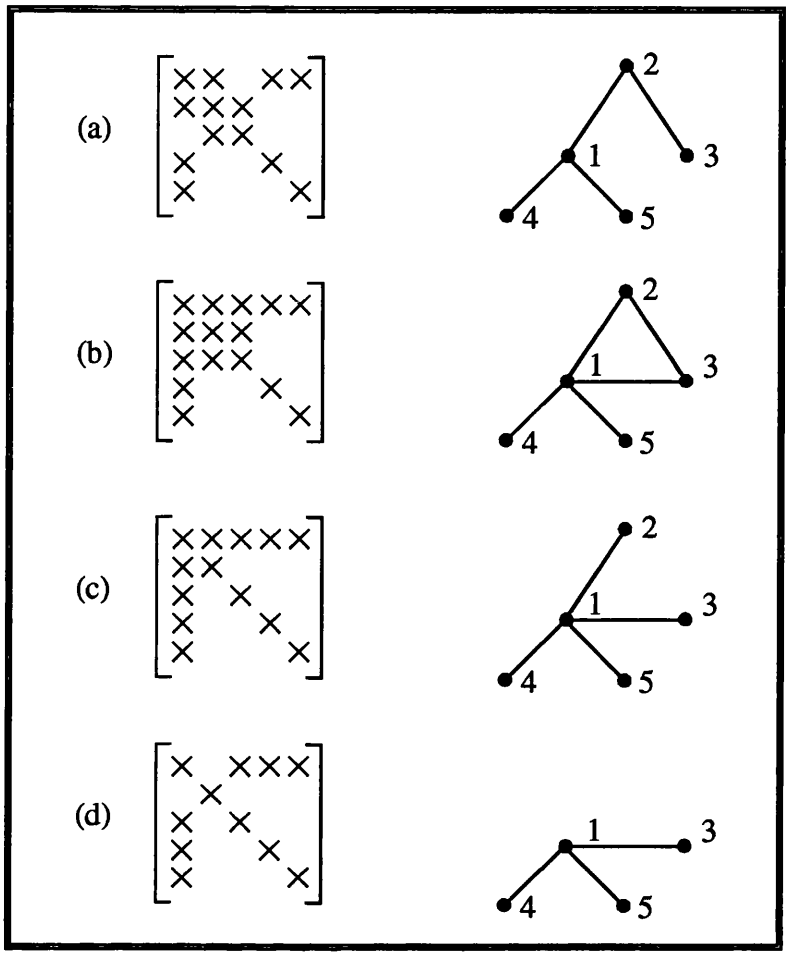


Figure A1.3(a) A tree graph and matrix
(b) Result after row/column subtraction on (a)
(c) First step in the elimination of node 2 from (a)
(d) Completing the elimination of node 2 from (a)

A1.2. Elimination Graphs

Sometimes, for example in the minimum degree algorithm, it is desirable to eliminate a node from the graph. The example below shows how node 2 in Figure A1.3(a) can be eliminated.

In a row or column subtraction, a multiple of row 2 or column 2 could be subtracted from row and column 3. Using the tree graph in Figure A1.3(a); if this multiple is chosen so that the subtraction produces a zero value in positions a_{23} and a_{32}, then the new graph is as pictured in Figure A1.3(c). Using the same technique to remove entries a_{12} and a_{21}, the resulting graph is shown in Figure A1.3(d). Note that the transformation

from Figure A1.3(a) to (c) to (d), removes all of node 2's edges and creates a new edge between node 2's original connecting nodes, i.e. 3 & 1. The removal of this node from the graph is termed **elimination** and the resulting subgraph is called the **elimination graph**. Note that in this case only 2 row and column subtractions were performed to eliminate node 3 as there were 2 off-diagonal zeros in row/column 3. If node 1 was to be eliminated from Figure A1.3(a), new edges would be created between nodes 2 & 4, 2 & 5 and 4 & 5.

A2. Matrix Conditioning

The subject of this appendix is matrix conditioning and how matrix eigenvalues are related to the conditioning. The effects of matrix conditioning on digital computer matrix calculations are examined.

A2.1. Eigenvalues and Matrix Conditioning

An eigenvalue and corresponding eigenvector of an $n \times n$ matrix **A**, satisfy the property that the eigenvector multiplied by the matrix yields a vector proportional to itself,

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{A2.1}$$

The constant of proportionality, λ_i is the eigenvalue and can also be viewed as one of the n roots of **A**'s characteristic polynomial. Matrix **A** will thus have n eigenvalues, of which not all may be distinct, since it is possible to have multiple roots, e.g. $\lambda_1 = \lambda_2$. If **A** has fewer than n distinct eigenvalues, the matrix is singular. The rank of a square matrix is equal to the number of non-zero eigenvalues. In general, the eigenvalues of a real matrix are complex numbers. However, when **A** is symmetric, all the eigenvalues of **A** are real. When all the eigenvalues of a symmetric matrix are strictly positive, the matrix is positive-definite.

The condition number is a norm-dependent property and is commonly based on the 2-norm. When **A** is a square and symmetric matrix, the 2-norm condition number, κ_2, is the ratio between the maximum and minimum eigenvalues,

$$\kappa_2(\mathbf{A}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|} \tag{A2.2}$$

For any general matrix **A**, the condition number is the ratio between the first and last singular values,

$$\kappa_2(\mathbf{A}) = \frac{|\sigma_1|}{|\sigma_n|} \tag{2.3}$$

A matrix is said to be ill-conditioned or ill-posed if the condition number is very large.

A2.2. Effects of Matrix Ill-conditioning

Stuart[95] shows that if **E** is a matrix of small errors, the relative error in $\mathbf{A}+\mathbf{E}$ (square matrices) may be magnified by as much as $\kappa(\mathbf{A})$ in finding $(\mathbf{A}+\mathbf{E})^{-1}$. Clearly, if the condition number is large, then the inverse of a matrix is sensitive to small perturbations in it and the calculation of the inverse can be seriously affected by ill-conditioning. Stuart[95] puts this in context with the following breakdown.

Suppose a t-digit approximation **B** to **A** is available, that is, $\|\mathbf{A}-\mathbf{B}\|/\|\mathbf{A}\| \cong 10^{-t}$. If $\kappa(\mathbf{A}) = 10^p$, then the relative error in $\mathbf{B}^{-1}$ is approximately 10^{p-t}. When the elements of $\mathbf{A}^{-1}$ are about equal in magnitude, they will be accurate to about t-p digits. This applies directly to solving a set of equations $\mathbf{Ax} = \mathbf{b}$ using finite-precision arithmetic. The relative error in **x** can then be $\kappa(\mathbf{A})$ times the relative error in **A** and **b**.

The Sun Sparc 10/51 used to obtain the results in this thesis has a machine precision of 1.1×10^{-16}, that is, any double precision number can be stored accurately to just under t=16 digits. Thus any matrix which has conditioning in excess of $\sim10^{16}$ is going to have a meaningless inverse when calculated on this computer. Any matrix with conditioning of this size and larger is then singular to machine precision.

A3. Transfer Impedance Matrix Calculation

This appendix summarises Hua's[91] method to find the transfer impedance matrix, **R** for any given resistivity distribution so that the optimal currents can be found. The diagonal-based currents are used as the basis current patterns.

Diagonal based currents use the same chosen electrode for sinking current. The source electrode is changed for each possible combination. For e electrodes, the diagonal basis patterns are then e-dimensioned vectors:

$$\begin{aligned}\mathbf{C}_1 &= \left[1/\sqrt{2}, 0, 0 \cdots 0, -1/\sqrt{2}\right]^{\mathrm{T}} \\ \mathbf{C}_2 &= \left[0, 1/\sqrt{2}, 0 \cdots 0, -1/\sqrt{2}\right]^{\mathrm{T}} \\ \mathbf{C}_{e-1} &= \left[0, 0, 0, \cdots 1/\sqrt{2}, -1/\sqrt{2}\right]^{\mathrm{T}}\end{aligned} \tag{A3.1}$$

Note that all the elements in each basis sum to zero, satisfying Kirchoff's law. These current patterns, $\mathbf{C}_i, i = 1, e-1$ are applied and the voltage responses collected from all electrodes, $\mathbf{V}_i, i = 1, e-1$. Using the current bases in equation (A3.1) and Ohm's Law, $\mathbf{V}_i = \mathbf{R}\mathbf{C}_i$, a matrix of all the measured voltages is formed and simplified as

$$\mathbf{V} = \left[\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_{e-1}\right] = \frac{1}{\sqrt{2}}\left[\mathbf{R}_1 - \mathbf{R}_e, \mathbf{R}_2 - \mathbf{R}_e, \cdots, \mathbf{R}_{e-1} - \mathbf{R}_e\right] \tag{A3.2}$$

where $\mathbf{R}_i$ is an e-dimensioned column vector in the $e \times e$ matrix $\mathbf{R}$. $\mathbf{R}$ is an indefinite matrix and so all the elements in a row sum to a constant. The same voltages arise for any given current pattern no matter what this constant is. Zero is chosen for simplification,

$$\sum_{j=1}^{e} \mathbf{R}_j = \mathbf{0} \tag{A3.3}$$

Summing the elements in each row in equation (A3.2) gives the following vector

$$\sum_{j=1}^{e-1} \mathbf{V}_j = \frac{1}{\sqrt{2}}\left(\left(\mathbf{R}_1 + \mathbf{R}_2 + \cdots + \mathbf{R}_{e-1}\right) - (e-1)\mathbf{R}_e\right) \tag{A3.4}$$

Substituting equation (A3.3) into (A3.4) and re-arranging, the expression for the last column in $\mathbf{R}$ in terms of measured voltages is

$$\mathbf{R}_e = -\frac{\sqrt{2}}{e}\sum_{j=1}^{e-1} \mathbf{V}_j \tag{A3.5}$$

The first i=e-1 columns of $\mathbf{R}$ are easily found by re-arranging equation (A3.2)

$$\mathbf{R}_i = \sqrt{2}\mathbf{V}_i + \mathbf{R}_e \tag{A3.6}$$

Each vector of measured voltages, $\mathbf{V}_j$ uses the grounded node as the voltage reference. Note that because of the zero constant chosen in equation (A3.3), each vector of measured voltages must be adjusted to a new voltage reference value, by imposing the constraint

$$\sum_{i=1}^{e} \mathbf{V}_{ij} = 0 \tag{A3.7}$$

A4. Finite Element Method

This part of the appendix derives the element equations using variational principles. The FEM assembly process and the element stiffness matrix for the bilinear element is given.

To find the voltages at the various nodes using the mathematical approach, the governing equation (3.1) is assigned, for convenience, by different variables and is rewritten as

$$\nabla \cdot \rho^{-1} \nabla \phi = 0 \tag{A4.1}$$

where ϕ is the field variable or voltage and ρ is the known resistivity of the mat.

The description of the field problem is not complete until boundary conditions are specified i.e. the governing equation must be solved subject to additional constraints imposed on the boundary. The boundary, in this case, is the electrodes. The relevant boundary conditions are

$$\phi = \bar{\phi} \text{ on } \partial A_{\phi} \tag{A4.2}$$

and

$$\rho^{-1} \frac{\partial \phi}{\partial \nu} = \bar{\phi}_{\nu} \text{ on } \partial A_{\nu} \tag{A4.3}$$

where $\bar{\phi}$ and $\bar{\phi}_{\nu}$ are the voltage and current density at the boundary region; ∂A_{ϕ} and ∂A_{ν} are the parts of the boundary where ϕ and $\partial\phi/\partial\nu$ are respectively chosen. The boundary of the resistive mat or domain A is denoted by $\partial A = \partial A_{\phi} + \partial A_{\nu}$. The term

$\partial\phi/\partial\nu$ is the gradient of the potential distribution normal to the boundary, ν denotes the unit outward normal vector on the boundary. Both the voltages and current densities are given functions along the boundary.

Equation (A4.2) is known as the Dirichlet[96] or rigid[97] boundary conditions, i.e. boundary conditions requiring the potential to take on prescribed values, and equation (A4.3) is known as the natural[97] boundary condition. If a variational approach is used to construct an approximate solution, the rigid boundary conditions must be satisfied in order to obtain a unique solution[97].

A4.1. Variational principles

Equation (A4.1) is Poisson's equation and represents a differential formulation. The problem is to integrate this differential equation subject to the boundary conditions given in equations (A4.2) and (A4.3). This differential formulation for the continuum problem has an equivalent variational formulation. In the classical variational formulation, the problem is to find the unknown function that minimises or maximises a functional subject to the same given boundary conditions. Before the variational approach can be used to derive the finite element equations, a variational statement for the continuum problem must be obtained. A functional $\pi(\phi)$ is sought such that its first variation with respect to ϕ vanishes. Variational principles yielding functionals for physical processes are found through classical variational calculus or by well established theorems in physics. The variational formulation for the differential equation and boundary conditions in hand, is obtained from Huebner[44] and Tong[97] as follows

$$\pi(\phi)=\tfrac{1}{2}\int_A \rho^{-1}\left[\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2\right]dA-\int_{\partial A_\sigma}\bar{\phi}_\nu\phi\,ds \qquad \text{(A4.4)}$$

where $dA=dxdy$ and ds is a line segment along the boundary of A. The resistivity, ρ is taken to be isotropic. The variational functional π, is the potential energy.

A4.2. Element Equations

The elements used are rectangular in shape, or more specifically square-shaped and are represented by the one in Figure 4.2(a). This element has 4 nodes and the value of ϕ at the 4 nodes are denoted by ϕ_1, ϕ_2, ϕ_3 and ϕ_4, where the subscripts are the local node

numbers for the element. The potential is chosen to vary linearly along all 4 sides and can be shown[97] to be represented by

$$\phi(x,y)=\sum_{i=1}^{4}\phi_i f_i(x,y) \tag{A4.5}$$

where the function f_i is called the interpolation function and has the value 1 at the ith node and the value 0 at the other three nodes. The interpolation function is an approximate solution within the element for the given boundary conditions that $\phi=\phi_i$ at node i and varies linearly along the boundaries.

In the finite element method, the potential energy in equation (A4.4) can be summed over each element,

$$\pi(\phi)=\sum_{h=1}^{m}\pi_h \tag{A4.6}$$

where m is the total number of elements in the domain, and

$$\pi_h=\tfrac{1}{2}\int_{A_h}\rho^{-1}\left[\left(\frac{\partial\phi}{\partial x}\right)^2+\left(\frac{\partial\phi}{\partial y}\right)^2\right]dA-\int_{(\partial A_v)_h}\bar{\phi}_v\phi\,ds \tag{A4.7}$$

with A_h equal to the area of the hth element and $(\partial A_v)_h$ is the portion of the boundary of the hth element into which the current is injected or extracted.

To obtain the element equations, differentiate equation (A4.5) with respect to x and y and substitute the result, as well as equation (A4.5), into equation (A4.7). The element equations are thus

$$\pi_h=\tfrac{1}{2}\boldsymbol{\phi}_h^{\mathrm{T}}\mathbf{k}_h\boldsymbol{\phi}_h-\boldsymbol{\phi}_h^{\mathrm{T}}\mathbf{Q}_h \tag{A4.8}$$

where

$$[\mathbf{Q}_i]_h=\int_{(\partial A_v)_h}\bar{\phi}_v f_i(x,y)\,ds \tag{A4.9}$$

and $\mathbf{k}_h$ is called the element stiffness matrix and is defined as

$$\left[\mathbf{k}_{ij}\right]_h = \int_{A_h} \rho_h^{-1}\left[\frac{\partial f_i}{\partial x}\frac{\partial f_i}{\partial x}+\frac{\partial f_i}{\partial y}\frac{\partial f_i}{\partial y}\right]dA \tag{A4.10}$$

If each element is regarded as homogeneous, ρ is a constant and equation (A4.10) becomes

$$\left[\mathbf{k}_{ij}\right]_h = \rho_h^{-1}\int_{A_h}\left[\frac{\partial f_i}{\partial x}\frac{\partial f_i}{\partial x}+\frac{\partial f_i}{\partial y}\frac{\partial f_i}{\partial y}\right]dA \tag{A4.11}$$

In equation (A4.9), $\mathbf{Q}$ represents the applied currents. Unless the ith node is the position of an injected or extracted current, $(\mathbf{Q}_i)_h$ is zero, otherwise it is the value of the injected/extracted current.

A4.3. Assembling the Element Equations

Substituting equation (A4.8) into equation (A4.6) yields

$$\pi = \sum_{h=1}^{m}\left(\tfrac{1}{2}\boldsymbol{\phi}_h^{\mathrm{T}}\mathbf{k}_h\boldsymbol{\phi}_h - \boldsymbol{\phi}_h^{\mathrm{T}}\mathbf{Q}_h\right) \tag{A4.12}$$

With algebraic manipulation, this equation can be written as

$$\pi = \tfrac{1}{2}\boldsymbol{\phi}^{\mathrm{T}}\mathbf{K}\boldsymbol{\phi} - \boldsymbol{\phi}^{\mathrm{T}}\mathbf{Q} \tag{A4.13}$$

The matrix, $\mathbf{K}$, is the master stiffness matrix of the entire domain and is symmetric and positive semi-definite[97]. When the model grows with an increasing number of elements, not all nodes are connected to each other, rather only to their adjacent nodes. This leads to $\mathbf{K}$ being very sparse.

This manipulation procedure is known as the assembling of the element matrices into the master matrix. In other words it is forming a larger model, shown in Figure 4.1, from the separate elements in Figure 4.2(a). Each element has its own local node numbers, as in Figure 4.2(a). In the assembly procedure, each node is given a unique global number, for example, those of Figure 4.1. The same value of ϕ is used for all the elements at the common nodes so that the approximate ϕ is continuous over the entire domain. In the variational formulation, ϕ must be at least continuous because π will not be

mathematically defined where ϕ is discontinuous[97]. Assembling the element stiffness matrix into the master stiffness matrix is a standard procedure:

$$\mathbf{K} = \sum_{h=1}^{m} [\mathbf{K}]_h = \sum_{h=1}^{m} \sum_{i=1}^{4} \sum_{j=1}^{4} [\mathbf{k}_{IJ}]_h \tag{A4.14}$$

where m is the total number of bilinear elements. The element stiffness matrix, $[\mathbf{K}]_h$ has elements $[\mathbf{k}_{IJ}]_h$. These matrix elements are those in equation (A4.11) where I & J are respectively the global node numbers of the local node numbers i & j.

When the element stiffness equation (A4.11) is assembled for a bilinear element, the element stiffness matrix is

$$\mathbf{k}_h = \frac{1}{6\rho_h} \begin{bmatrix} 4 & -1 & 2 & -1 \\ & 4 & -1 & 2 \\ & \text{sym} & 4 & -1 \\ & & & 4 \end{bmatrix} \tag{A4.15}$$

The solution of ϕ is obtained from the variational principle by applying[44]

$$\frac{\partial \pi}{\partial \phi} = 0 \tag{A4.16}$$

to equation A(4.13) to give

$$\mathbf{K}\phi = \mathbf{Q} \tag{A4.17}$$

This is the same as equation (4.1) in Section 4 of the thesis.

Appendix B Published Research

1) WS Fulton and RT Lipczynski. *Electrical Impedance Tomography Applied to Body-Support Interface Pressure Measurement.* Abstract from presentation to Biological Engineering Society 33rd Annual Scientific Meeting, University of Bath, 13-15 September 1993.

2) WS Fulton and RT Lipczynski. *Body-Support Pressure Measurement Using Electrical Impedance Tomography.* Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, USA, 28-31 October 1993, Volume 15 pages 98-99.

3) WS Fulton and RT Lipczynski. *Optimising the time to solution in electrical impedance tomography.* IEE Proceedings - Science, Measurement and Technology, Volume 142, Number 6, November 1995, pages 433-441.

Electrical Impedance Tomography Applied to Body-Support Interface Pressure Measurement

W S Fulton & R T Lipczynski
School of Electrical Engineering
University of Bath

Abstract

The research focuses on the implementation of a body-support interface pressure measurement system. The intended applications for interface pressure measurement are in the medical field to study the pressure distribution of seated wheelchair patients or bed-ridden patients. In certain circumstances, these patients will be prone to pressure sore development. The pressure measurement data that will be produced will enable medical physicians to study the pressure distributions that lead to pressure sores. The data will also enable the assessment of the different prophylactic devices that alleviate and spread pressure, as a comprehensive evaluation of these has not yet been undertaken. A system could also be placed in each hospital ward to estimate the probability of pressure sore formation in admitted patients.

This proposed pressure measurement system is unique in that it measures pressure over a continuum. Current pressure measurement systems are limited in that they only give pressure readings at certain points. The proposed system needs a ½ mm thin piezo-resistive mat placed between the body and supporting medium. The mat exhibits a change in resistivity at points at which a load is applied. Electrical Impedance Tomography (EIT) is a medical imaging technique that can be applied to measure the resistivity distribution. The conventional applications for Electrical Impedance Tomography are in imaging body organs, bones and tissues and access to the centre of the resistivity distribution is prohibitive. This access enables the large effect that data measurement errors have on image accuracy to be overcome. The largest errors introduced by three-dimensional current flow are removed and those introduced by variable electrode positioning are reduced. Reconstruction results showing these improvements will be shown.

BODY-SUPPORT PRESSURE MEASUREMENT USING ELECTRICAL IMPEDANCE TOMOGRAPHY

W.S. Fulton and R.T. Lipczynski

School of Electrical Engineering
The University of Bath
Claverton Down, Bath, BA2 7AY, U.K.

ABSTRACT

This paper presents a computer simulation of a system to measure the interface pressure distribution between a patient and support surface. This is required for the analysis and study of pressure distributions of seated wheelchair or bed-ridden patients who are prone to pressure sore development. The proposed system uses Electrical Impedance Tomography techniques to obtain the conductivity distribution of a pressure sensitive resistive mat placed between the patient and support surface. Good results are obtained when utilising access to the centre of the mat for data measurement.

INTRODUCTION

Pressure sores are an unpleasant infliction, yet they are commonplace in most hospitals and nursing homes. In the U.S.A. the medical costs associated with pressure sores were estimated to be US$2 billion[1]. The only means of reducing the incidence of pressure sores is by the prevention of them. Improved methods of treatment would at the most only have a marginal effect[2]. Regular patient turning is the ideal solution to the prevention of pressure sore formation, but many underfunded and understaffed wards simply do not have the facilities to devote time to this task. Exton-Smith[2] concludes that it is now unrealistic to expect nurses to practise regular two-hourly turning, except in rare instances, and scarce nursing skills should now be deployed for the recognition of high-risk patients who require the use of specialised equipment to take over these tasks.

There are many different prophylactic devices that alleviate and spread pressure, but a comprehensive evaluation of these is yet to be undertaken. Barbenel[3] explains that the information obtained by measuring pressure does not completely characterise the body-support interface, but, even with this limitation, pressure measurements have a most clearly established clinical usefulness and that both pressure and time for which it acts, are important variables in assessing the potential for tissue damage by any supporting surface. Other requirements for pressure distribution data is in the design of wheelchair cushions and in the analysis of gait. Presently available clinical pressure measurement equipment can only produce point readings, whose information content is somewhat limited and whose sensors interfere with the true pressure distribution. The proposed system enables a pressure distribution over an entire area to be realised.

PRESSURE MEASUREMENT METHOD

The measurement system requires a flexible pressure sensing mat placed between the body and supporting surface. The mat must have known transverse resistivity characteristics under an applied normal load. Piezo-resistive materials satisfy these requirements and are fabricated from either carbon fibres or carbon loaded elastomers. Data is collected from the mat via simple electronics and then passed on to a personal computer for processing. The electronics is required to inject currents through point electrodes connected to the mat and to measure the resulting voltages at other electrodes. Thus, by reconstructing the two-dimensional resistivity distribution, a profile of the applied pressure is obtained.

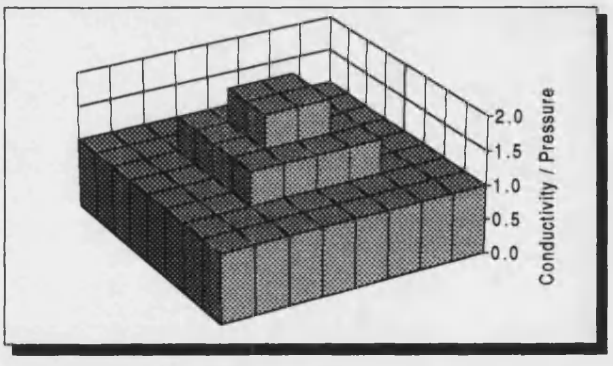


Figure 1. The FEM mesh of the test distribution

The resistivity distribution is discretised by the Finite Element Method (FEM), as shown in Figure 1. Electrical Impedance Tomography (EIT) is an imaging technique which is then used to reconstruct the resistivity profile. The problem of resistivity reconstruction is non-linear and different EIT reconstruction methods have been developed[4]. Our chosen method is that which uses a Gauss-Newton reconstruction, first developed for EIT

imaging by Yorkey[5]. The Gauss-Newton algorithm has been programmed to reconstruct the resistivity/pressure profile from simulated voltage readings obtained from the FEM. To simulate practical measurement errors, the true voltage signals were contaminated with random noise. Two electrode configurations have been used. The first utilises 32 electrodes connected to the edge of the mat. The second uses the same number of electrodes, 28 connected to the edge and 4 connected to the central area of the mat. Four different current patterns inject current between two diagonally opposite electrodes. In the second electrode configuration, one central electrode replaces an edge electrode.

RESULTS

A range of resistivity distributions were reconstructed, one of which is shown in Figure 1. The algorithm converges to the true solution with no simulated voltage error. With an added error of 0.03% of the average voltage, the reconstruction errors for the two different electrode configurations are shown in Figures 2 and 3. Similar results were obtained for a range of added error levels. In the practical environment, the pressure would be in the central area. Various other profiles of central pressure perturbations produced similar results.

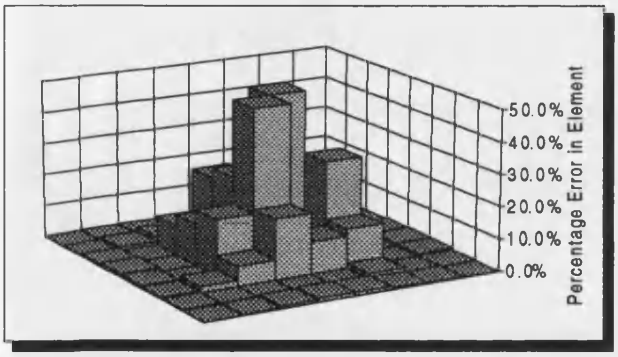


Figure 2. Reconstruction errors using the edge electrode configuration

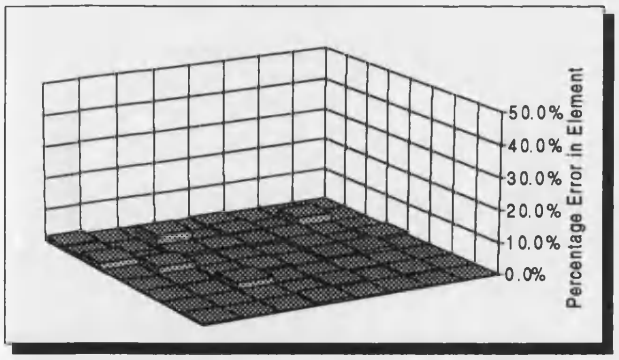


Figure 3. Reconstruction errors using the internal electrode configuration

DISCUSSION

The benefit of moving just four of the current injection and voltage measurement nodes towards the centre of the mat is quite clear. The limited spatial resolution in the central region[6] leads to the high error in these elements when using edge electrodes only. The Hessian matrix approximation in the Gauss-Newton algorithm is ill-conditioned. This causes a low tolerance to measurement error and matrix inversion to be inaccurate. The Hessian matrix 1-norm condition number for the profiles shown is 7×10^8 when utilising the edge configuration and 9×10^6 for the internal configuration. The use of internal electrodes for current injection and voltage measurement improves the matrix conditioning and consequently the tolerance to error in the voltage readings. The Gauss-Newton method is still susceptible to errors, but can be improved further through the use of regularisation[4]. The source of errors would be due to electrode positioning, inaccuracies in the finite element model and those in the electrical system. The errors are, however, expected to be low for a fine FEM mesh and accurately placed electrodes.

In a practical environment, the pressure profiles are expected to be with the load in the central area. Under these conditions, we have shown that by placing electrodes in the central area, the matrix conditioning improved, leading to good reconstruction results.

ACKNOWLEDGEMENTS

This work is funded by the British Science and Engineering Research Council. The authors would like to thank L.J. Grant, Head of the Wessex Regional Health Authority, for inspiring the research in this area.

REFERENCES

[1] S.L. Garber, "The role of technology in tissue management," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 8, pp. 1835-1836, 1986.

[2] N. Exton-Smith, "The patients not for turning," *Nursing Times*, vol. 83, October 21, pp. 42-44, 1987.

[3] J.C. Barbenel, "Measurement of interface pressures," in *Pressure Sores*, J.C. Barbenel, C.D. Forbes and G.D.O. Lowe, Eds, London, MacMillan, 1983, pp. 67-78.

[4] P. Hua and E.J. Woo, "Reconstruction Algorithms," in *Electrical Impedance Tomography*, J.G. Webster, Eds. Bristol: Adam Hilger, 1990, pp. 97-137.

[5] T.J. Yorkey, J.G. Webster and W.J. Tompkins, "Comparing reconstruction algorithms for Electrical Impedance Tomography," *IEEE Trans. Biomed. Eng.*, vol. BME-34, pp. 843-852, 1987.

[6] A.D. Seagar, D.C. Barber and B.H. Brown, "Theoretical limits to sensitivity and resolution in impedance imaging," *Clin. Phys. Physiol. Meas.*, vol. 8, Suppl A, pp. 13-31, 1987.

Optimising the time to solution in electrical impedance tomography

W.S. Fulton
R.T. Lipczynski

Indexing terms: Electrical impedence tomography, Sparse matrix methods, Nested dissection

Abstract: Electrical impedance tomography (EIT) is an inverse imaging technique used to reconstruct resistive distributions. Some EIT inverse methods, such as the Gauss–Newton method, use the finite element method in the forward problem to solve for the Jacobian matrix. The paper details methods to improve the solution time in the forward problem and accurate Jacobian calculation. The first suggestion is to keep the same node grounded for all projection angles by floating the current source. Details of an optimised Jacobian formation are also presented including full exploitation by sparse vector methods. The application of EIT to interface pressure distribution measurements makes use of a pressure sensitive resistive mat that can be cut to any shape or size. This flexibility can be taken advantage of and a square or rectangular domain comprising square bilinear elements is chosen. The sparse matrix nested dissection node renumbering algorithm then gives further speed advantages and also enables best optimisation with sparse vector techniques.

1 Introduction

Electrical impedance tomography is a technique which is able to construct an image of the distribution of electrical impedance, within an electrically conducting object, from a knowledge of the voltage and current conditions at various points on the object. Electrical impedance tomography has applications in a number of fields. Typical geophysics applications are for obtaining core sample data, geophysical probing for detection of minerals at the earth's surface and in borehole scanning. Process tomography uses EIT to image conducting fluids within vats or flowing within pipes [1]. In the medical field, body organs, bones and other tissues have different impedances enabling EIT to image them [2]. A fourth application of EIT is in the measurement of interface pressure. Interface pressure measurement has many applications, e.g. monitoring and studying pressure sore development in wheelchair-bound or bed-ridden patients. In pressure measurement a thin, flexible pressure-sensitive mat exhibiting known pressure-resistivity characteristics is used [3,4]. The pressure distribution is simply mapped from the mat resistivity that the EIT technique reconstructs.

The reconstruction of the domain's resistivity distribution is broken into a forward and an inverse problem. The forward problem involves finding voltage distributions given the resistivity distribution and certain boundary conditions. The inverse problem solves for the resisitive distribution given measured voltages and current density distributions. To solve the inverse problem (the defined problem), a numerical solution is employed, using iterative solutions of the forward problem. The Gauss–Newton or modified Newton–Raphson method [5] is chosen for the inverse solution, primarily for its accurate reconstructions when the system errors, i.e. electrode placement errors, finite element method modelling errors, electronic noise etc. are low. Interface pressure measurement, a primary target of this paper, has lower system errors than most EIT applications as the electrodes can be accurately placed and the currents are not three dimensional. Errors in a pressure EIT application also have less negative effect on the Gauss–Newton reconstruction as electrodes may be utilised throughout the domain [6].

The EIT problem is ill-conditioned. This leads to small errors in the Jacobian destabilising the Gauss–Newton method. Regularisation [7] of the Gauss–Newton method is used to offset these noise induced instabilities, thereby reducing the reconstruction error. The iterative regularised Gauss–Newton method is defined by a resistivity vector ρ^k updated at any iteration k

$$\rho^{k+1} = \rho^k + \Delta\rho^k \tag{1}$$

$$\Delta\rho^k = -\left[[\mathbf{f}']^T\mathbf{f}' + \lambda\mathbf{S}\right]^{-1}[\mathbf{f}']^T[\mathbf{f} - \mathbf{v}_o] \tag{2}$$

where $\mathbf{f}$ is a function mapping the m-dimensioned resistivity distribution vector into an n-dimensioned voltage vector. The voltage observation vector, $\mathbf{v}_o$ is obtained from voltage measurements at electrodes connected to the pressure mat or domain. The matrix, $\mathbf{f}'$ is called the Jacobian matrix, $\mathbf{S}$ is the regularisation matrix and λ is a parameter dictating the amount of regularisation used. When λ is zero, no regularisation is introduced and eqn. 2 then defines the plain Gauss–Newton method. The finite element method (FEM) solves the forward problem. The solution of the finite element equations and the solution of the Jacobian matrix are closely related. The efficient solution of these equations is desirable as this has a dramatic decrease on the time to solution of the iterative Gauss–Newton method, making real-time pressure scanning more feasible.

IEE Proceedings online no. 19952083

Paper received 30th November 1994

The authors are with the School of Electronic and Electrical Engineering, University of Bath, Claverton Down, Bath BA2 7AY, UK

2 Solving the finite element equations

This Section explains the sparse solution of the finite element equations as a six stage process.

The finite element equations are solved for each projection angle

$$\mathbf{Yv} = \mathbf{c} \tag{3}$$

where $\mathbf{Y}$ is the $r \times r$ symmetric master stiffness matrix evaluated at ρ^k and $\mathbf{v}$ and $\mathbf{c}$ are r-dimensioned vectors, which hold the nodal voltages and currents, respectively, for any one projection angle. The finite element mesh contains r nodes and the finite element equation solves the nodal voltages for all r nodes. The vector $\mathbf{f}$ is obtained by extracting just the electrode voltages for each projection angle. The well-known Cholesky method is the standard approach for solving the nodal voltages in eqn. 3. The Cholesky decomposition factors $\mathbf{Y}$ into the form $\mathbf{LL}^T$, where $\mathbf{L}$ is lower triangular. eqn. 3 becomes

$$\mathbf{LL}^T\mathbf{v} = \mathbf{c} \tag{4}$$

The following triangular systems are then solved by forward and backward substitution

$$\mathbf{Lx} = \mathbf{c} \tag{5}$$

$$\mathbf{L}^T\mathbf{v} = \mathbf{x} \tag{6}$$

The master stiffness matrix, $\mathbf{Y}$ is also called the coefficient matrix and is a symmetric positive definite matrix. It is also sparse and consequently eqns. 4–6 are suitable for sparse matrix solution. Good overviews of sparse matrix techniques have been given by Duff *et al.* [8] and George and Liu [9]. The $\mathbf{Y}$ matrix structure results from the node numbering in the FEM mesh. The $\mathbf{Y}$ matrix structure resulting from a natural (row by row) ordering on a 20×20 or 400 element grid is shown in Fig. 2. In this and other similar diagrams, nonzeros are represented by black squares. The white space indicates where the zero elements lie in the matrix. The matrix structure is evident from the positioning and the number of these nonzeros. Fig. 3 shows the factored matrix, $\mathbf{L}$, which has added nonzero entries (fill-in) where there were zero entries in the original $\mathbf{Y}$ matrix and so it has a slightly modified structure from $\mathbf{Y}$. In this particular example, fill-in of the entire area between the diagonal and the $\mathbf{Y}$ matrix nonzeros has occurred. This need not always happen and judicious choice of the mesh node numbering can reduce the percentage fill-in. A reduction in fill-in is desirable as it leads to fewer floating point operations when using sparse techniques to solve the equations.

The sparse solution of the finite element equations is broken into the six steps shown in Fig. 1. Note that for each repeated solution of the right hand side vector, neither the master stiffness matrix assembly nor factorisation is repeated in any one Gauss–Newton iteration as the structure of the master stiffness matrix is the same for the same finite element mesh. The first three steps are performed only once for any number of iterations. The six steps are briefly considered individually.

2.1 Pseudo-assembly

The finite element assembly procedure produces the r×r master stiffness matrix structure which must be known for some of the node renumbering algorithms to be employed. The finite element assembly procedure is used solely to find the positions of the offdiagonal nonzero entries in the master stiffness matrix. This pseudo-assembly produces a one-dimensional array

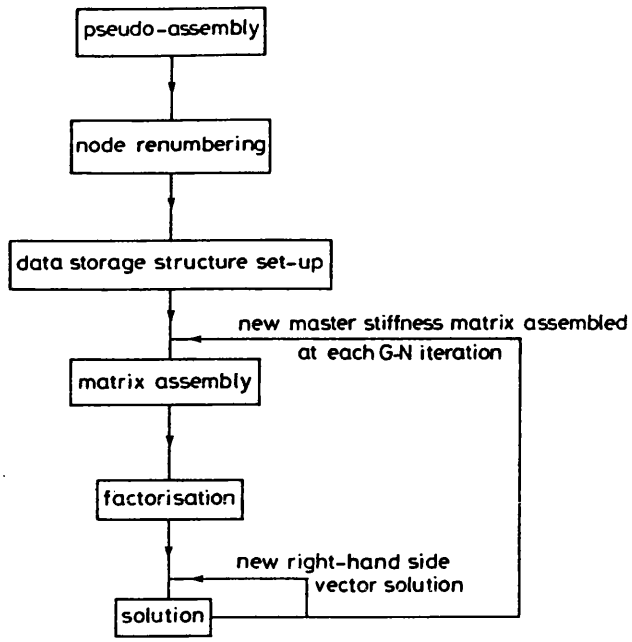


Fig. 1 *Steps to solution of sparse finite element equations in the Gauss–Newton method*

which has r segments, each corresponding to a column in $\mathbf{Y}$. Each segment stores a list of the row numbers at which there is a nonzero in the $\mathbf{Y}$ matrix. A pointer array which points to the first nonzero element of each column in the one-dimensional array is also needed. Various matrix entries are referenced more than once during the assembly procedure, so the pseudo-assembly implements a search ensuring that entries are not duplicated. The resulting one-dimensional array is needed for the node renumbering schemes used.

2.2 Node renumbering

Numerous node numbering schemes exist and the choice of a suitable numbering scheme is an essential part of the sparse matrix technique as it determines the numerical operations count. This, in turn, determines the time taken for the factorisation stage and also affects the sparse vector operations count in the forward and backward substitution stages. Renumbering nodes from the original ordering is equivalent to interchanging rows and columns of $\mathbf{Y}$ through the use of a permutation matrix, $\mathbf{P}$. The permutation matrix is simply the identity matrix with permutated rows and columns. eqn. 3 strictly needs rewriting as

$$(\mathbf{PYP}^T)\mathbf{Pv} = \mathbf{Pc} \tag{7}$$

and the Cholesky decomposition is applied to these re-ordered equations. For simplification, $\mathbf{P}$ is not used in this paper and $\mathbf{Y}$, $\mathbf{v}$ and $\mathbf{c}$ refer to the re-ordered matrix/vector. The node numbering schemes are considered in detail in Section 4.

2.3 Data storage structure set-up

Once the matrix re-ordering is chosen, the sparse data storage structures must be set up for storage of the re-ordered coefficient matrix and the factored matrix, $\mathbf{L}$. The storage scheme is linked to the type of sparse factorisation and solution method used as follows.

(1) Envelope methods produce a band structure with predictable fill-in which facilitates the use of simple data structures for matrix storage. The variable band structure consists of rows of different lengths within the

band, as shown in Figs. 2 and 3. The area within the band is also known as the envelope. The data storage structure used is the envelope storage scheme [9] and stores all the elements within the envelope.

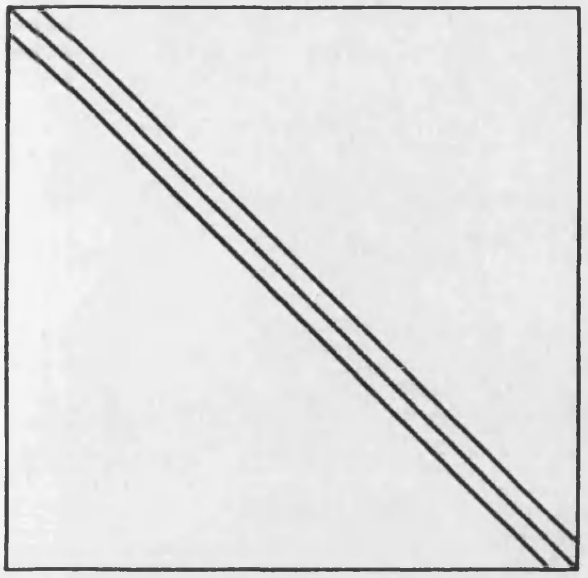


Fig. 2 **Y** *matrix structure for 400 element grid with natural ordering*
3721 nonzeros, 1.9% sparsity

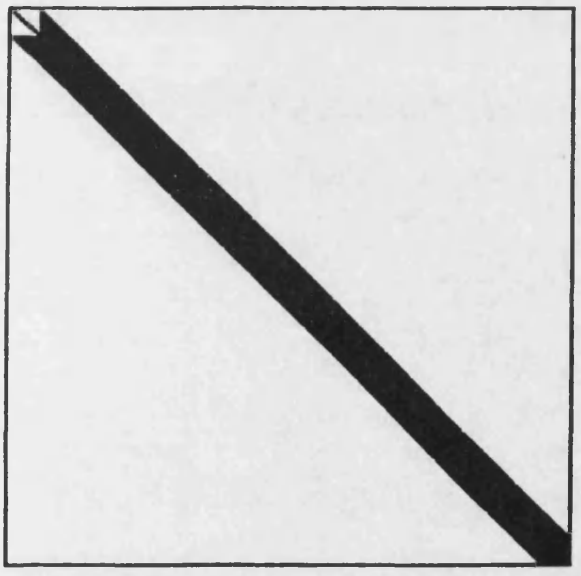


Fig. 3 **L/L**^{T} *matrix structure for 400 element grid with natural ordering*
18 921 nonzeros, 15 200 (7.8%) fill-in

(2) The more general minimum degree and nested dissection sparse matrix orderings, which will be considered, produce a more scattered distribution of nonzero elements (Figs. 4–9). A more complicated data storage structure is needed to access these elements. The solvers used require a data storage structure especially adapted to symmetric sparse factorisation and solution. The compressed sparse scheme [9] uses a one-dimensional array to store the coefficient and factored matrices, along with the necessary pointer and element positioning vectors. These more general sparse matrix orderings do not give the same factored matrix fill-in predictability that is associated with envelope structures. To find the positions of the fill-ins, a process known as symbolic factorisation [9] must be performed to find the zero/nonzero structure of the factored matrix. The compressed sparse data storage scheme is set up during the symbolic factorisation stage.

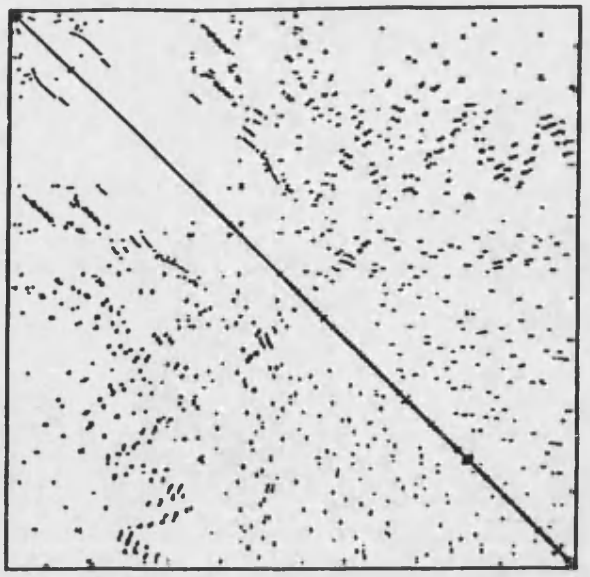


Fig. 4 **Y** *matrix structure for 400 element grid with minimum degree ordering*
3721 nonzeros, 1.9% sparsity

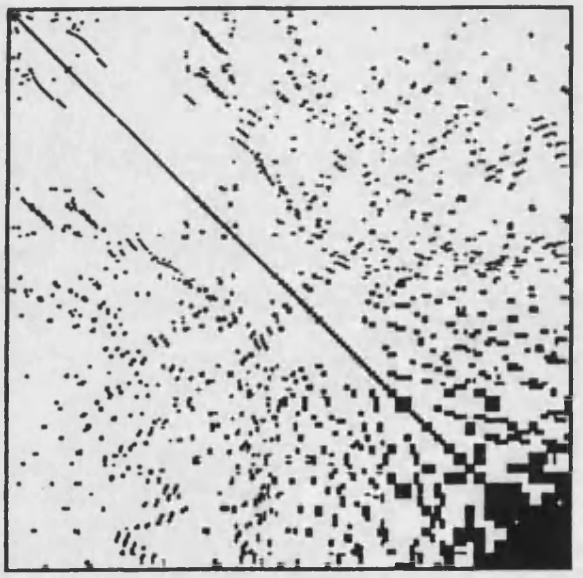


Fig. 5 **L/L**^{T} *matrix structure for the 400 element grid with minimum degree ordering*
14075 nonzeros, 10354 (5.3%) fill-in

2.4 Matrix assembly

The coefficient matrix is formed from the FEM assembly process. Only the lower triangle of **Y** is stored in the appropriate one-dimensional data arrays, so only the elements in the lower triangle must be assembled, as follows:

$$\mathbf{Y} = \sum_{h=1}^{m} \sum_{i=1}^{4} \sum_{j=1}^{i} \left[\mathbf{Y}_{\max(I,J)\,\min(I,J)}\right]_h = \sum_{h=1}^{m} \sum_{i=1}^{4} \sum_{j=1}^{i} \left[\mathbf{k}_{ij}\right]_h \tag{8}$$

where m is the total number of bilinear elements in the mesh, $\mathbf{k}_{ij}$ is an element in the FEM element stiffness matrix, and I and J are the global node numbers for element hs local node numbers i and j, respectively.

Standard FEM techniques require the boundary conditions to be implemented after **Y** has been assembled. The vector **c** in eqn. 3 contains the natural boundary conditions, that is, the currents applied to the mat. This

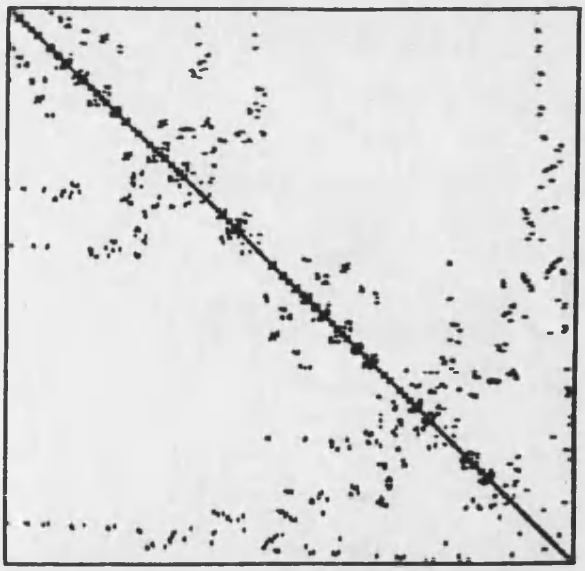


Fig. 6 Y *matrix structure for 400 element grid using irregular nested dissection ordering*
3721 nonzeros, 1.9% sparsity

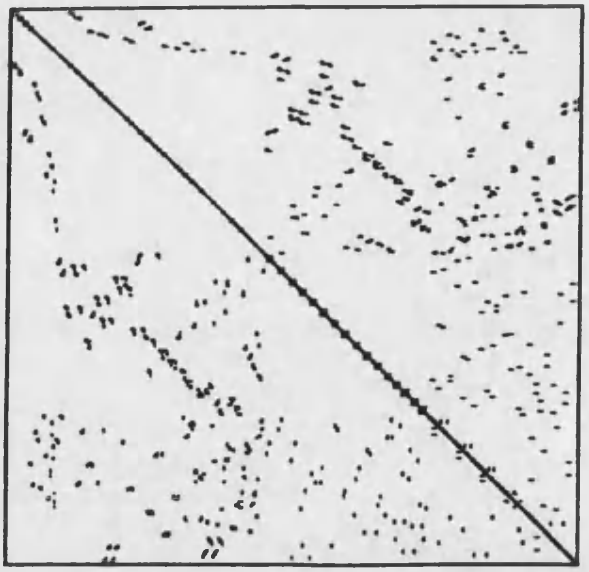


Fig. 8 Y *matrix structure for 400 element grid using regular nested dissection ordering*
3721 nonzeros, 1.9% sparsity

Fig. 7 L/L^T *matrix structure for 400 element grid using irregular nested dissection ordering*
13461 nonzeros, 9740 (5.0%) fill-in

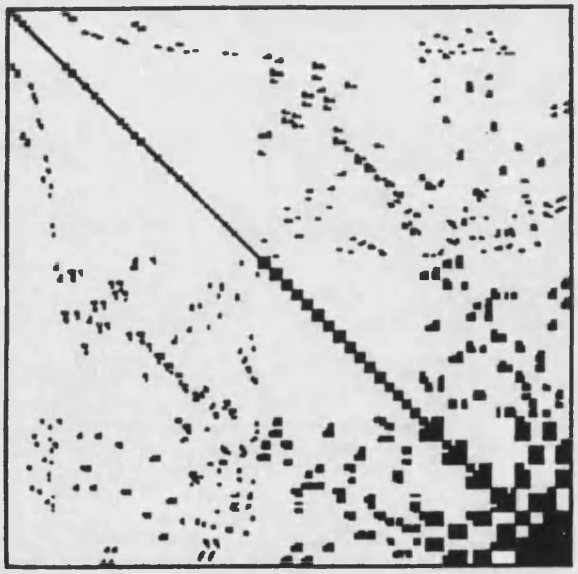


Fig. 9 L/L^T *matrix structure for 400 element grid using regular nested dissection ordering*
12735 nonzeros, 9014 (4.6%) fill-in

vector thus holds the current in amperes at row numbers which correspond to the electrodes and the rest of the vector contains zero elements. **Y** is nonsingular, so a reference node is supplied by grounding a node k. This is the equivalent of implementing rigid boundary conditions, $v_k = 0$ into eqn. 3.

2.5 Factorisation

After the boundary conditions have been implemented, **Y** is a positive definite symmetric matrix, which is a necessary condition for the Cholesky factorisation. Matrix **Y** is factored into the form $\mathbf{LL}^T$ using the Cholesky factorisation algorithm. Efficient implementation of the numerical factorisation algorithm is dependent on the storage scheme used. Appropriate library routines [9] optimise the numerical factorisations according to the data storage structure scheme used.

2.6 Solution

Eqn. 3 is solved once for each projection angle. The different **v** vectors are each solved separately through the forward and backward substitutions of eqns. 5 and 6. Section 5 shows that **c** is sparse in some situations and the sparse vector methods examined in this Section enable further computational savings to be made during the solution stage.

3 Solving for the Jacobian matrix

This Section suggests projection angles which allow the Jacobian to be calculated once per Gauss–Newton iteration. An optimised solution of the Jacobian matrix that eliminates redundant calculations is also presented. The Jacobian is an $np \times m$ partitioned matrix, the structure of which is shown in Fig. 10. The number of elements in the finite element mesh is denoted by m and the number of projection angles used is denoted by p. Each partition corresponds to a projection angle. The Jacobian matrix is comprised of mp n-dimensioned vectors, $\mathbf{f}'_{ij}$ as shown in Fig. 10.

Woo [10] has derived a method to solve for the Jaco-

bian matrix through the use of an $rp \times m$ partitioned matrix $\mathbf{Q}$ by differentiating the finite element equation eqn. 3 for each projection angle. Matrix $\mathbf{Q}$ has the same structure as the Jacobian, but is comprised of mp r-dimensioned vectors, $\mathbf{q}_{ij}$ calculated by forward and backward substitution using

$$\mathbf{Y}\mathbf{q}_{ij} = -\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{v}_i \tag{9}$$

where $\mathbf{v}_i$ is the $\mathbf{v}$ column vector for the ith projection angle. Only n elements in the r-dimensioned $\mathbf{q}_{ij}$ vector are required and so the redundant elements are removed to form the relevant $\mathbf{f}'_{ij}$ vector in the Jacobian shown in Fig. 10.

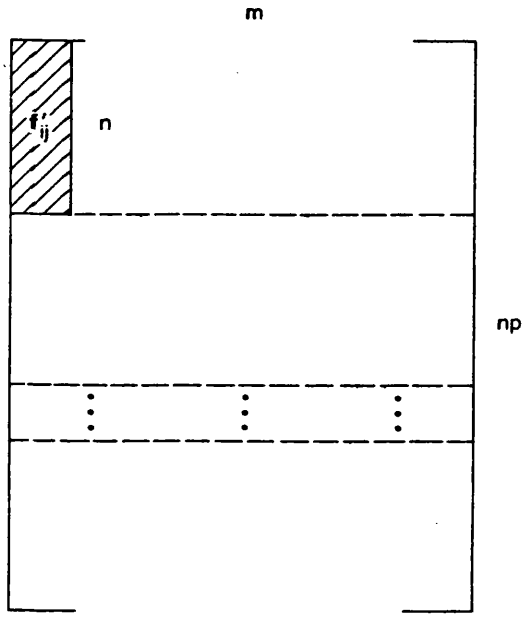


Fig. 10 *Jacobian matrix structure showing one f'_{ij} vector and p partitions*

A more efficient method to obtain the Jacobian is to form the $\mathbf{f}'_{ij}$ vector directly, thereby avoiding the calculation of the unwanted elements in $\mathbf{q}_{ij}$. In eqn. 9, only n rows out of the r calculated are needed. Substantial savings are possible as $n \ll r$ when using grids with large numbers of elements. Eqn. 9 is modified to calculate $\mathbf{f}'_{ij}$ directly

$$\mathbf{f}'_{ij} = -\mathbf{G}\frac{\partial \mathbf{Y}}{\partial \rho_j}\mathbf{v}_i \tag{10}$$

where $\mathbf{G}$ is an $n \times r$ matrix and contains just the required n rows of $\mathbf{Y}^{-1}$ to form $\mathbf{f}'_{ij}$. As $\mathbf{Y}^{-1}$ is a symmetric matrix, $\mathbf{G}$ can be formed from either the columns or the rows of $\mathbf{Y}^{-1}$. Columns can be obtained very efficiently through forward and backward substitution of the equation $\mathbf{Y}\mathbf{Y}^{-1} = \mathbf{I}$, where $\mathbf{I}$ is the $r \times r$ identity matrix (see Section 5). Only the required columns in $\mathbf{Y}^{-1}$ to form $\mathbf{G}$ are solved for.

The factored matrix of $\mathbf{Y}$ is obtained during the finite element equation calculation and so is re-used here for the solution of $\mathbf{G}$ for all projection angles. Re-use for all projections is only possible if the grounded or reference node is kept in the same position for each projection, i.e. the rigid boundary conditions are kept constant. Two types of commonly applied projections are optimal current patterns [11] and point current injection. Optimal current patterns usually keep the reference node fixed and apply current to all other electrodes. In point current injection, each projection angle involves the current source being attached to two different electrodes. EIT researchers use one of these electrodes as the grounded node. Alternatively, if the grounded node is left in the same place for each projection angle for voltage measurement, a floating current source can be attached to two other nodes. In this way all the current injected by one node can be extracted by the second and $\mathbf{Y}^{-1}$ remains the same for each projection angle. More unique information is obtained for greater numbers of projection angles. Keeping one node grounded for all projection angles decreases the amount of unique information gained per additional projection angle. If maximum unique information is required, then $p=e$ projections must be used, where e is the total number of electrodes. The accuracy in the reconstruction is dependent on the amount of unique information in the Jacobian, but often $e/2$ or fewer projections are used. It was found that this reduction in unique information had a minimal effect on the quality of the reconstruction as regularisation is used. Usually with point current injection, the rows in $\mathbf{G}$ that relate to the two current carrying electrodes are not used. Then $n = e - 2$ electrode rows are used so that the voltage drop, due to the electrode impedance on the two current carrying electrodes, does not introduce any measurement error.

Yorkey *et al.* [5] noted that the derivative, $\partial \mathbf{Y}/\partial \rho_j$ is sparse which allows for further computational savings to be made by only performing the nonzero multiplications when multiplying by $\mathbf{v}_i$ in eqns. 9 and 10. The bilinear interpolation function used produces an element stiffness matrix which has 16 elements and thus only contributes 16 nonzero elements to both the global impedance matrix and its derivative. This is the case no matter how large the dimensions of either $\mathbf{Y}$ or $\partial \mathbf{Y}/\partial \rho_j$.

4 Sparse matrix re-orderings

The sparse matrix technique, in effect, finds the permutation matrix in eqn. 7. Use of the permutation matrix corresponds to relabelling of the node numbers in the finite element mesh and gives rise to the matrix structure. An optimal matrix structure is sought as this will reduce the number of fill-ins and subsequent operation counts during the factorisation and solution stages. Two commonly used approaches to node renumbering are the envelope orderings and the minimum degree ordering. These are briefly examined along with a more detailed explanation of another approach using nested dissection.

4.1 Envelope orderings

Matrices can be ordered so that all the nonzeros lie within an envelope around the main diagonal. During factorisation, fill-in occurs within this envelope and so envelope algorithms operate entirely within this area, ignoring all the zeros outside it. The natural or row-by-row ordering of the mesh produces the narrow band structure of Figs. 2 and 3 for the square 400 element grid. Algorithms that automatically order matrices into a band structure exist and the best of these is known as the Gibbs–Poole–Stockmeyer algorithm. This algorithm attempts to reduce the size of the envelope, but fails to decrease the size that the natural ordering produces.

4.2 The minimum degree ordering

The minimum degree algorithm is also known as the Tinney scheme 2 ordering and was first published by Tinney and Walker [12] in 1967. Since then, the algorithm has been improved slightly and more efficiently implemented [13]. The algorithm is specially tailored for a wide range of symmetric matrices and is a popular ordering for many sparse matrix problems.

The minimum degree algorithm attempts to re-order the matrix for minimum fill-in during the factorisation stage of the Cholesky decomposition. However, it does not always produce an ordering that gives the minimum amount of fill-in [8]. A symmetric matrix can be represented by an undirected graph in graph theory. It can be verified [8] that when the graph has no loops (a tree graph), the minimum degree ordering introduces zero fill-in. The graph for the bilinear element used is shown in Fig. 11. Zero fill-in is not possible as this graph structure has loops and the minimum degree algorithm cannot guarantee an ordering that will produce minimum fill-in. Woo [10] recommended using the minimum degree ordering but this was on a circular EIT domain comprised of triangular elements. Its performance needs analysing on a square domain with square bilinear elements. The resulting renumbered 400 element mesh produces the coefficient and factored matrices shown in Figs. 4 and 5.

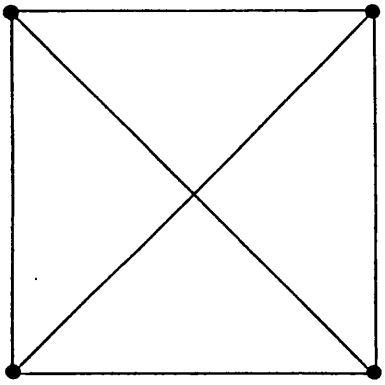


Fig. 11 *Undirected graph of single bilinear finite element*

The minimum degree algorithm used has no tie-breaking strategy, that is, when two or more nodes have equal minimum degree, no strategy is used to choose one for numbering next. The choice of node is then entirely dependent on the initial ordering. Different tie-breaking strategies induce different orderings and so the amount of fill-in and the success of the minimum degree algorithm depends on the tie-breaking strategy used. George and Liu [13] found that a natural initial ordering on a square grid was best for minimum fill-in and so this is the one used.

4.3 Nested dissection orderings

Nested dissection is an ordering algorithm which can be adapted to meshes comprised of both regular and irregular shaped finite elements. The regular square finite element mesh is the relevant case, but an irregular algorithm may also be applied to a regular problem. Duff *et al.* [14] published a nested dissection algorithm especially tailored for regular square finite element grids.

The nested dissection ordering algorithm is based on dividing up the finite element mesh piece by piece. The mesh is separated into four subsets by a cross-shaped group of dividing nodes. The process is repeated on the remaining nodes in each subset until each subset can no longer be subdivided. The dividing group of nodes, or dissection sets, in each subset are numbered last. The resulting four dissection sets for a 10x10 or 100 element grid are shown in Fig. 12, with a few sets shaded for clarification. The nested dissection algorithm numbers the nodes in set 1 first, followed by those in sets 2,3... etc. Methods for numbering the nodes within each dissecting set can be optimised and have been studied by Duff *et al.* [14]. He found the minimum degree ordering to be best. This node numbering applied to the 100 element mesh using the dissecting sets of Fig. 12 is shown in Fig. 13. The resulting nonzero matrix structure for the 400 element mesh is shown in Figs. 8 and 9.

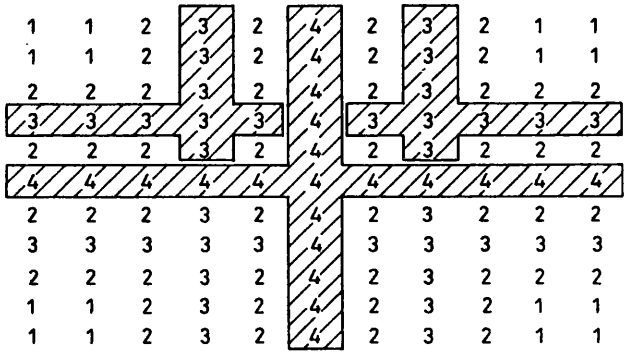


Fig. 12 *Regular nested disection sets for 100 element mesh 4th disection set and two of the 3rd dissection sets are shaded*

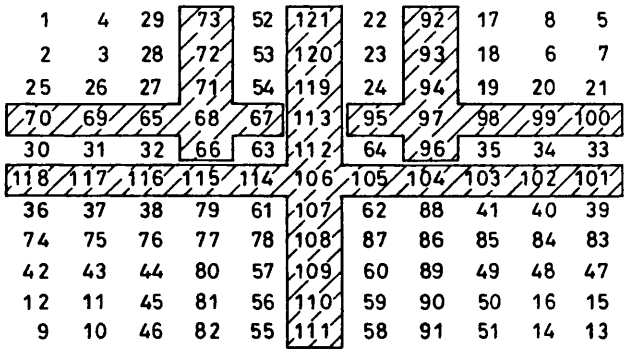


Fig. 13 *Regular nested dissection node numbering for 100 element mesh using same dissection sets as in Fig. 12*

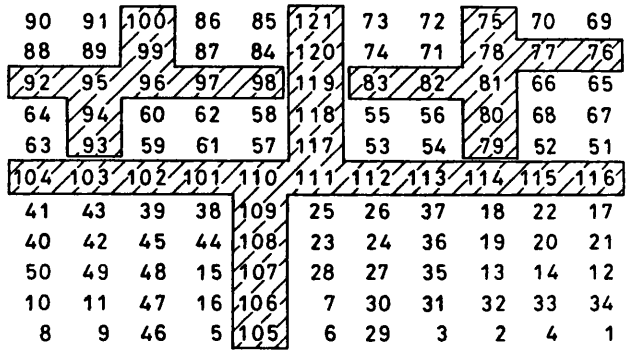


Fig. 14 *Irregular nested dissection node numbering for 100 element mesh produced by automatic algorithm; A few groups within the dissection sets are shaded to identify them*

George and Liu [15] developed an automatic general purpose nested dissection ordering algorithm which can also be applied to irregular finite element problems. The authors claim that it gives an efficient ordering for regular square finite element grids. The ordering produced for the 100 element mesh case is shown in Fig. 14. Comparisons with the regular mesh ordering

of Fig. 13 show that the dissecting sets are not quite the same. The consequences of these differences is examined in Section 6. The resulting matrix structures for the 400 element grid are shown in Figs. 6 and 7.

5 Sparse vector methods

The incentive for using sparse vector methods [16] is to improve on the sparse matrix method by further reducing the number of unnecessary floating point operations. This can be achieved in two instances: when the right hand side vector is sparse and/or only a subset of the solution vector needs to be known. In the first instance, forward substitution can be optimised by using a subset of the columns of the factored matrix, **L** and is known as fast forward (FF). Similarly in the second instance, a subset of the rows of $\mathbf{L}^T$ speeds up backward substitution and is known as fast back (FB). The ordered list of columns in **L** for FF, or rows in $\mathbf{L}^T$ for FB, which are absolutely necessary for solution, define what is known as the sparse vector path. The path length in FF and FB is the number of these necessary columns in the ordered list.

The paths for FF and FB are usually different and are, respectively, functions of the sparsity structure in the right-hand side and solution vectors. The paths are also functions of the sparsity structure in **L**, which, in turn, is a function of the **Y** matrix ordering strategy used. The time spent in the forward and backward substitution stages is small in comparison to the factorisation stage. Sparse vector methods are thus only worth using when many solutions using the same coefficient matrix are needed. Both the finite element voltage and Jacobian solution use the same coefficient matrices many times over and the following subsections examine whether FF and FB should be used.

5.1 Finite element voltage solution

Vector **c** can only contain nonzeros in positions that correspond to the electrode numbers used by the current source. In single current injection, current is injected into one electrode and extracted from another and so **c** has just two nonzeros. This is a prime candidate for using FF as **c** is highly sparse. When optimal currents are used, each and every electrode has an injected current. In this case, **c** is not particularly sparse. Additionally, the electrodes are not topologically clustered. Sparse vector methods (FF) are not considered in this case as these conditions [16] are not ideal for using them. The Jacobian matrix calculation needs all voltages on all nodes for the Jacobian matrix calculation so full backward substitution is necessary.

5.2 Jacobian matrix solution

Section 3 details an optimised method to solve for the Jacobian matrix. This method needs a total of e columns in $\mathbf{Y}^{-1}$ and all the elements of each column. The equation $\mathbf{Y}\mathbf{Y}^{-1} = \mathbf{I}$ is solved to obtain the rows in **G**. Each column vector in **I** is sparse with just one nonzero. Such a sparse vector is called a singleton by Tinney *et al.* [16] and is the best condition for solution optimisation by FF. As all elements in the solution vector are required, full backward substitution is necessary.

6 Results and analysis

This Section analyses the sparse matrix node numbering and sparse vector method combinations for minimising the solution time. A method must be chosen that is optimal for both the voltage solution and the inverse matrix solution used in the formation of the Jacobian. A second analysis is on the computational savings of the improved Jacobian matrix calculation.

6.1 Sparse solution results

The algorithms were programmed in Fortran 77, many of which were available as efficiently coded library routines [9]. The number of operations, i.e. the number of multiplications and divisions, are used to examine the algorithms' efficiencies. These are then tested on a superscalar computer, a Sun 10/41. The time spent in the factorisation and solution stages dominates the time spent in all six stages. The time spent in the remaining stages is negligible and thus is not shown.

Table 1 compares the number of operations executed during Cholesky factorisation. It shows that the nested dissection algorithms have superior operation counts and the algorithm specially tailored for the regular mesh (RND) is better than the general irregular algorithm (IND). The IND algorithm still outperforms the minimum degree (MD) algorithm on larger meshes. Operation count improvements are larger on larger meshes. The amount of fill-in has a direct effect on the operation count. This is apparent in the percentage fill-in shown with the matrix structures of Figs. 2–9.

Table 1: Operation count for Cholesky factorisation

m	25	64	100	256	400
Nat	990	4 496	9 580	50 336	113 560
MD	802	3 550	7 574	43 781	85 041
IND	936	4 232	8 546	35 929	71 667
RND	808	3 460	7 005	32 282	64 987

Nat natural or row by row ordering
MD minimum degree ordering
IND ordering from the irregular/automatic nested dissection algorithm
RND regular mesh nested dissection ordering
number of bilinear elements in the square mesh is denoted by m

The solution stage produces similar results. The full backward substitution and full forward substitution/singleton FF counts can be seen in Table 2. The operation count reduction by FF is largely due to a reduced path length. The average singleton FF path length for calculating all columns in $\mathbf{Y}^{-1}$ is shown in Table 3 along with the full forward substitution path length for comparison. The natural ordering is not particularly suited to FF as once the first right hand side nonzero is encountered, the path is full from then on. The more scattered nonzero distributions in the minimum degree and nested dissection ordered matrices lead to path lengths which are much shorter than those of the full path. These orderings are more suitable for sparse vector methods. The nested dissection algorithms have shorter path lengths than the minimum degree algorithm and are thus particularly suited to FF.

The RND ordering produces the lowest operation count for both FF and full forward substitution. This is reflected in timing comparisons for one Gauss–Newton iteration. An example comparing the two larger grid sizes is shown in Table 4. The times shown use the full forward substitution for the voltage calculation, as would be the case with optimal currents. The three smaller meshes showed similar timing trends, but

Table 2: Average operation counts comparing full forward substitution and singleton FF

	25		64		100		256		400	
m	Full	FF	Full	FF	Full	FF	Full	FF	Full	FF
Nat	492	365	1 602	1 195	2 882	2 153	10 370	7 763	19 362	14 503
MD	436	288	1 366	888	2 428	1 502	8 590	5 391	14 516	8 661
IND	472	296	1 506	903	2 606	1 543	8 084	4 671	13 902	7 947
RND	438	281	1 358	832	2 348	1 425	7 618	4 471	13 176	7 649

Operation counts include full backward substitution
see Table 1 for key

Table 3: Forward substitution path lengths

m	25	64	100	256	400
Full	36	81	121	289	441
Nat	18	41	61	145	221
MD	12	22	26	52	59
IND	10	17	23	39	50
RND	11	18	24	41	54

Nat, MD, IND, RND and *m* are defined in Table 1 and show singleton FF path lengths
Full is the full forward substitution path length

their modelling inaccuracies are not ideal for accurate reconstruction.

The time to solution is also dependent on the data storage scheme used by the sparse solver. This is evident from the lower factorisation time of the envelope solver when compared to the MD general sparse solver, even though the operation count is higher. However, as Table 4 indicates, the majority of time spent in one iteration is in the solution stage. The RND ordering has shorter path lengths and this overrides the advantage of the envelope solver's fast access to its simpler storage structures.

Table 4: Timing comparisons for different orderings using appropriate solvers

	m	256	400
Envelope solver (Nat)	Factorisation	12.30	35.10
	Full solution	2.33	4.75
	FF solution	1.59	3.28
	Total	100	216
General sparse solver (MD)	Factorisation	15.70	40.60
	Full solution	2.01	3.80
	FF solution	1.18	2.03
	Total	86	166
General sparse solver (IND)	Factorisation	13.60	35.70
	Full solution	1.96	3.59
	FF solution	1.00	1.84
	Total	77	152
General sparse solver (RND)	Factorisation	12.00	30.50
	Full solution	1.86	3.40
	FF solution	0.92	1.66
	Total	71	138

e = total number of electrodes
p = number of projection angles
Times are in milliseconds on a Sun 10/41. Total time is for one Gauss–Newton iteration, i.e.
total = factorisation + full solution * p + FF solution * e, where e=32 and p=16
see Table 1 for key

6.2 Computational savings of the improved Jacobian matrix calculation

The number of floating point operations show the computational savings for the proposed Jacobian calculation of Section 3. The improved method based on eqn. 10 has $16mp(n + 1) + nz_1$ operations, whereas the method based on eqn. 9 requires $mp(16 + z_2)$. Each variable z_1 and z_2 contains the sum of forward and backward substitution operations in the appropriate sparse matrix/vector method. Full forward substitution must be used for z_2, whereas FF can be used for z_1. The comparisons are more tangible with an example. The example uses the largest mesh for good spatial resolution, i.e. m=400, p=16, n=e=32. The regular nested dissection has been shown to be the optimal sparse matrix and vector approach, so the RND counts in Table 2 are used for z_1 and z_2. The floating point operation's ratio for eqns. 9 and 10 is then 23:1. The reduction in redundant calculations is clearly demonstrated by this example. The ratio is larger for meshes with many elements (large m) as then m and z are the dominant counts. Yorkey *et al.* [5] proposed a method to approximate the Jacobian matrix which they estimated to take $15mn$ operations. However, this approximation is less efficient than the method based on eqn. 9 when optimal current patterns are used [17]. This Jacobian approximation is not suited to optimal current patterns as the method is based on point current injection. The improved method suggested in this paper is thus most suitable for use with optimal current injection.

The more elements there are in the mesh, the larger is the problem and the greater is the effect of using the suggested sparse techniques and removal of redundant calculations. This is evident from the combined time spent calculating the voltage vector and the Jacobian as a percentage of the total time spent in one Gauss–Newton iteration. For the example set-up described in the previous paragraph, this figure is 11% and for the smaller 25 element problem, 70%. Previously, this time completely dominated one iteration but has now been reduced to a minor part for the more accurate larger problems.

6.3 Further discussion

Many different ordering strategies have been compared with each other in the sparse matrix literature. The fastest solutions depend to a large degree on the matrix sizes, the original ordering and the degree number of each node [14] (type of element used). Both the way the algorithms are coded and the type of microprocessor used have some effect on the solution time. Consequently, it is difficult to foresee which combination of sparse vector and sparse matrix method is best without physically testing the various algorithms. Woo [10] examined envelope ordering methods and three Tinney–Walker [12] schemes in his EIT system. He suggested using the minimum degree ordering for a circular finite element mesh comprised of triangular elements. The irregular nested dissection algorithm is better than the minimum degree algorithm on the

square domain, but could be used for any mesh as it is independent on the mesh structure and is more suited to sparse vector methods.

Better tie-breaking strategies might be able to improve the minimum degree algorithm. Tie-breaking strategies are a topic of recent research [18]. However, the nested dissection ordering can be viewed as a minimum degree ordering with a perfect tie-breaking strategy when applied to regular grids [13]. Another reason for the success of the nested dissection algorithm is that it is a global ordering strategy as the most important decisions are made on the entire matrix, whereas the minimum degree algorithm is a local ordering strategy [8]. The solution methods considered in this paper are direct methods especially adapted to symmetric sparse matrices. Iterative methods are alternative approaches to a solution, but were not considered as they are best for single solutions, are better for larger problems and are more efficient when a good starting vector is available. The work described here is independent of whether or not regularisation is used, as regularisation merely involves including matrix S in the updating eqn. 2. In fact, any inverse EIT method that requires a Jacobian matrix or uses the FEM to solve for the nodal voltages can use the optimising techniques described.

7 Conclusions

In real-time applications, a reduction in reconstruction time enables better resolution in the reconstructed distribution as more elements may be used in the FEM mesh, for any given time limit. Optimisations in the forward problem solution and the calculation of the Jacobian matrix have been presented. A method which removes redundant Jacobian matrix element calculations has been put forward and is most appropriate for use with optimal current patterns. This method and the finite element equations used in the forward problem are suitable for sparse matrix and sparse vector optimisation. The regular nested dissection algorithm has been shown to be the best node renumbering strategy for a regular rectangular mesh comprised of square bilinear elements. Additionally, the nested dissection ordering was shown to produce singleton path lengths shorter than those of the minimum degree ordering. The suitability of a mesh comprised of square bilinear elements for optimisation by sparse methods is one factor in choosing them to model the mat in our pressure scanner. Grounding one node for all projection angles allows the master stiffness matrix inverse to be calculated just once per iteration. Previously, the time spent in the forward problem and Jacobian matrix calculation completely dominated the reconstruction of the pressure distribution. This paper's suggestions have reduced it to a minor part of the Gauss–Newton inverse problem solution time.

8 Acknowledgments

The authors would like to thank the Science and Engineering Research Council for funding the work in this field. They are grateful for the sparse matrix display routines written by Chris Murphy. Thanks are also due to Professor Eastham and the School of Electronic and Electrical Engineering for supporting the research.

9 References

1 BECK,M.S., CAMPOGRANDE,E.C., MORRIS,M., WILLIAMS,R.A., and WATERFALL,R.C.: 'Tomographic techniques for process design and operation' (Computational Mechanics Publications, Southampton, 1993)
2 DIJKSTRA,A.M., BROWN,B.H., LEATHARD,A.D., HARRIS,N.D., BARBER,D.C., and EDBROOKE,D.L.: 'Clinical applications of electrical impedance tomography', *J. Med. Eng. Technol.*, 1993, **17**, pp. 89–98
3 KNIGHT,R.A.: 'The use of EIT techniques to measure interface pressure'. PhD thesis, University of Bath, Bath, UK, 1991
4 BASARAB-HORWARTH,I., and FARAJ,M.: 'A prototype distributed pressure sensor utilising electrical impedance tomography; initial results'. *in* BECK,M.S., CAMPOGRANDE,E.C., MORRIS,M., WILLIAMS,R.A., and WATERFALL,R.C.: (Eds.): 'Tomographic techniques for process design and operation' (Computational Mechanics Publications, Southampton, 1993), pp. 277–283
5 YORKEY,T.J., WEBSTER,J.G., and TOMKINS,W.J.: 'Comparing reconstruction algorithms for electrical impedance tomography', *IEEE Trans.*, 1987, **BME-34**, pp. 843–852
6 FULTON,W.S., and LIPCZYNSKI,R.T.: 'Body-support pressure measurement using electrical impedance tomography', Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1993, Vol. 15, pp. 98–99
7 HUA,P., WEBSTER,J.G., and TOMKINS,W.J.: 'A regularised electrical impedance tomography algorithm', *Clin. Phys. Physiol. Meas.*, 1988, **9**, (Suppl A), pp. 137–141
8 DUFF,I.S., ERISMAN,A.M., and REID,J.K.: 'Direct methods for sparse matrices' (Oxford University Press, New York, 1986)
9 GEORGE,A., and LIU,J.W.H.: 'Computer solutions of large sparse positive definite systems' (Prentice Hall, New Jersey, 1981)
10 WOO,E.J.: 'Computational complexity', *in* WEBSTER,J.G., (Ed.): 'Electrical impedance tomography' (Adam Hilger, Bristol, 1990)
11 GISSER,D.G., ISAACSON,D. and NEWELL,J.C.: 'Current topics in impedance imaging', *Clin. Phys. Physiol. Meas.*, 1987, **8**, (Suppl A), pp. 39–46
12 TINNEY,W.F., and WALKER,J.W.: 'Direct solutions of sparse network equations by optimally ordered triangular factorization', *Proc. IEEE*, 1967, **55**, pp. 1801–1809
13 GEORGE,A., and LIU,J.W.H.: 'The evolution of the minimum degree ordering algorithm', *SIAM Rev.*, 1989, **31**, pp. 1–19
14 DUFF,I.S., ERISMAN,A.M., and REID,J.K.: 'On George's nested dissection method', *SIAM J. Numer. Anal.*, 1976, **13**, pp. 686–695
15 GEORGE,A., and LIU,J.W.H.: 'An automatic nested dissection algorithm for irregular finite element problems', *SIAM J. Numer. Anal.*, 1978, **15**, pp. 1053–1069
16 TINNEY,W.F., BRANDWAJN,V., and CHAN,S.M.: 'Sparse vector methods', *IEEE Trans.*, 1985, **PAS-104**, pp. 295–301
17 YORKEY,T.J.: 'Electrical impedance tomography with piecewise polynomial conductivities', *J. Comput. Phys.*, 1990, **91**, pp. 344–360
18 GOMEZ,A., and FRANQUELO,L.G.: 'Node ordering for sparse vector method improvement', *IEEE Trans.*, 1988, **PWRS-3**, pp. 73–79